NATIONAL QUALITY FORUM

+ + + + +

BEHAVIORAL HEALTH PHASE 3
STANDING COMMITTEE MEETING

+ + + + +

THURSDAY
OCTOBER 2, 2014

+ + + + +

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, Medical Director, CDC, National Center for Chronic Disease Prevention and Health Promotion
HAROLD PINCUS, MD, Director of Quality and Outcomes Research, New York-Presbyterian Hospital, The University Hospital of Columbia and Cornell Universities
ROBERT ATKINS, M.D., MPH, Senior Medical Director, Aetna Medicaid
MADY CHALK, PhD, MSW, Director, Policy Center, Treatment Research Institute
DAVID EINZIG, MD, Medical Director of Child Psychiatry, Children's Hospital and Clinics Of Minnesota
JULIE GOLDSTEIN GRUMET, PhD, Director of Prevention and Practice, Education Development Center/Suicide Prevention Resource Center/National Action Alliance for Suicide Prevention

CONSTANCE HORGAN, Sc.D., Professor and Director, Institute for Behavioral Health, The Heller School for Social Policy and Management, Brandeis University
LISA JENSEN, DNP, APRN, Associate Director Workforce & Leadership, Office of Nursing Services, Veteran's Health Administration
DOLORES (DODI) KELLEHER, MS, DMH, Principal, D Kelleher Consulting
KRAIG KNUDSEN, PhD, Chief, Bureau of Research and Evaluation, Ohio Department of Mental Health and Addiction Services
MICHAEL LARDIERI, LCSW, Assistant Vice President, Strategic Program Development, North Shore-LIJ Department of Psychiatry
TAMI MARK, PhD, MBA, Vice President, Truven Health Analytics
RAQUEL MAZON JEFFERS, MPH, MIA, Director of Health Integration, The Nicholson Foundation
BERNADETTE MELNYK, PhD, RN, CPNP/PMHNP, FAANP, FNAP, FAAN, Associate Vice President for Health Promotion, University Chief Wellness Officer, Dean and Professor, College of Nursing, Professor of Pediatrics & Psychiatry, College of Medicine, The Ohio State University
LAURENCE MILLER, MD, Senior Psychiatrist, Arkansas Medicaid, Arkansas Medicaid
DAVID PATING, MD, Chief, Addiction Medicine, Kaiser Permanente
VANITA PINDOLIA, Pharm.D., VP, Ambulatory Clinical Pharmacy Programs, Henry Ford Health System/Health Alliance Plan
RHONDA ROBINSON BEALE, Medical Physician, Former Chief Medical Office at Optum Now Health Care Consultant, Health Care Consultant

HENA SIDDIQI, M.D., Medical Director, Broadlawn Manor Nursing and Rehabilitation
LISA SHEA, M.D., D.F.A.P.A., Deputy Medical Director, Quality and Regulation, Butler Hospital (Providence, RI)
JEFFERY SUSMAN, M.D., Dean, Northeast Ohio Medical University, Northeast Ohio Medical University
MICHAEL TRANGLE, MD, Associate Medical Director for Behavioral Health, HealthPartners
BONNIE ZIMA, MD, MPH, Professor in Residence, Child and Adolescent Psychiatry, UCLA Semel Institute for Neuroscience and Human Behavior
LESLIE ZUN, MD, MBA, Chair, Department of Emergency Medicine, Mount Sinai Hospital

NQF STAFF:

POONAM BAL, Project Analyst
HELEN BURSTIN, Chief Scientific Officer
LAURALEI DORIAN, Project Manager
ANGELA FRANKLIN, Senior Director

ALSO PRESENT:

STEVE DAVIS, M3 Information
KENDRA HANLEY, AMA PCPI
SARAH HUDSON SCHOLLE, NCQA
JUNQING LIU, NCQA
COLETTE PITZEN, Minnesota Community Measurement*
D.E.B. POTTER, AHRQ
SARAH SAMPSEL, MPH, Azul Quality Solutions, Consultant to NQF
COREY WALLER, Spectrum Health

* present by teleconference

A G E N D A

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

8:31 a.m.

Welcome, Recap of Day 1

CHAIR BRISS: So good morning and welcome.

MS. DORIAN: Good morning everyone, and welcome back to Day 2. It's a good sign that you're back, and we're excited to get started to review some exciting measures again today. Just before we get started on these measures, we wanted to remind you what happened yesterday. So Poonam will bring the slide up.

Okay. So the slide isn't working at the moment, but I can just remind everyone from my notes. So we reviewed -- the first measure we reviewed was the ADHD 1 from NCQA which passed.

The second one, which was the only one that didn't, was not recommended, was the pediatric symptom checklist and psychosocial functioning from Mass General, which went down

on reliability.

CHAIR BRISS: Although we made some recommendations about it being important, and some ways that it could be adjusted over time.

MS. DORIAN: Exactly, yes, and then here's the slide. And then so the Health Screening and Assessment, the NCQA measures were all recommended, but with the note the 2601, we had consensus not reached on validity.

So when that -- when the report goes out for public comment, we'll really solicit comments on that measure in particular, and then when this committee reconvenes for the post-comment call at the end of October, the end of November I think, we can discuss those comments that have been received.

CHAIR BRISS: And on that set too, there were a number of important -- there was a lot of important discussion about this

essentially being a step in the right direction, and I think there was a fair amount of consensus on the committee that we'd like to see more steps in the right direction, towards composites or stratifications.

MS. DORIAN: So were there any sort of overarching comments or questions about yesterday?

Tobacco, Alcohol and Substance Use

MS. DORIAN: Okay. I think we can jump right into our first measure then.

CHAIR BRISS: And just for planning purposes, so two things. We know already that we're going to start losing people this afternoon as they're running for planes, so it will be important to try to stay on time again today.

I appreciated everybody's hard work on being efficient yesterday, and we'll need to do a lot of that again, and we're going to propose for short term conflicts, we're going to do a slight modification of the

morning schedule.

So we'll start as proposed with 2597. We'll then move to the alcohol and tobacco measures, 2599 and 2600, after which we'll circle back and pick up 2605.

CHAIR PINCUS: So we're starting with the Substance Use Screening and Intervention Composite from the American Society of Addiction Medicine, and the measure developer is here.

MR. WALLER: Yes, the measure developer is here.

CHAIR PINCUS: Okay. So do you want to sort of tee this up? And this is a little bit different, and you want to talk about that Angela?

MS. FRANKLIN: This is a measure that's coming into our alternate pathway. We're looking at this eMeasure as a trial measure, meaning that our recommendations today will relate to whether the measure is to be used for testing, further testing, to

gather the data needed to determine the measure's reliability and validity.

So it won't be used for use in accountability applications. It will be for use for testing purposes, and you will have at your table a quick guide on how to evaluate this measure as a trial eMeasure. Just to give you a quick refresher, we'll be looking at this measure against the NQF criterion.

However, when it comes to scientific acceptability, we'll only be looking at the first criterion for reliability, looking to see whether the measure specifications are precise, and then we come to the validity criterion.

We'll only be looking at the first criterion again, which is the criterion as to whether the measure specifications are supported by the evidence presented for the measure.

Are there any questions about how we're looking at this measure today as a

committee?

DR. BURSTIN: Just one piece of context for this. This sort of comes out of the blue for you, so just a moment on this. So the idea is that we very much want to push on the idea of moving to really good new eMeasures.

Lots of new concepts. You know we can't keep getting out of claims, we can't keep cracking open medical records every time to do them.

But we also recognize that we're at the point right now a lot of the EHRs just aren't ready to test some of these new systems, and we found a lot of the developers are having a really hard time finding EHRs, three EHRs with which to test some of these measures, because they are new and important concepts.

So this has been our thinking, of at least allowing some of these measures to flow out there, get used, get tested and then

they'll just come back to us to bring actually to you guys likely, just to bring their testing results for reliability and validity.

It is at least sort of a stepwise progression, and at least pushing forward and allowing some of the more innovative concepts to move forward.

CHAIR BRISS: Carol, do you have a question?

CHAIR PINCUS: Oh yes, Tami.

MEMBER MARK: Yeah. Can you explain some reasons why we wouldn't want to go forward with testing, and then sort of thinking like it seems -- I'm not quite understanding the bar for voting against additional testing? It just seems like it would --

DR. BURSTIN: We're not asking you to vote against additional testing. We are expecting these measures to be tested. They're just coming to you today not yet tested. So they will not be endorsed. That's

another important distinction here.

What comes out of this designation by you is they are approved for trial use, trial use measures. Not endorsed, not endorsed until the testing results come back.

MEMBER MARK: So I guess why would we not approve something for trial use?

DR. BURSTIN: Because they may not pass the other criteria. It may not be -- it may not have any evidence in place. Among scientific acceptability, though, they're not yet tested for reliability and validity. You may not think those specifications are precise enough. There are other elements within scientific acceptability.

You would still want to look at feasibility. Very important for an eMeasure, and they're all required to have eMeasure feasibility testing, and then usability. So those other criteria are still in play, even if they have not yet had an opportunity to find three EHRs that could actually test it.

CHAIR BRISS: I would say there are also -- yes, this is hypothetical. But imagine a measure that was measuring something that you didn't think was important to measure, that was badly designed. You might ask them -- you might ask a developer to go back and redesign the measure before going to the time and expense of testing reliability and validity.

DR. BURSTIN: Right.

CHAIR PINCUS: Just one additional clarification. With regard to doing any kind of testing, are they expected to have done that, for example, with regard to feasibility, to have tested the feasibility? So it's really just the validity and the reliability testing.

DR. BURSTIN: Exactly, and NQF did some work with -- for ONC just about a year ago. The Office of National Coordinator came up with a specific testing approach for feasibility. So to at least ensure that these

measures are feasible as a starting point, before they start getting put into use. Another question?

MEMBER MAZON JEFFERS: I'm sorry. Could you just then reiterate which criteria we are expected to evaluate today for this particular measure, just so I'm clear?

DR. BURSTIN: Well, you will evaluate all four criteria. So none of them get pulled off the table, except within the scientific acceptability criteria. You will not vote on reliability and validity. We will only ask you to look at the remaining elements, which is really about precision of the specifications. Is that okay? Does that work?

CHAIR PINCUS: We do apply actually 2a1 and 2b1 for reliability and validity. It's the other reliability --

DR. BURSTIN: Just not the testing.

CHAIR PINCUS: Yes. Just not the

testing part of it. Okay. Any other questions? Rhonda and then Leslie.

MEMBER ROBINSON BEALE: Just a quick question. Given that this is an area that's using electronic health records, and even though there is a good penetration of electronic health records in the behavioral health space, it's still not as substantial as it is on the medical side.

Will I know, maybe -- it just has to be said. So I raised a question as to whether or not one will look at, as part of the feasibility, the penetration of electronic medical records in that particular area.

DR. BURSTIN: You know, it's a fair question, and I think some of the idea here is that we're just pushing them out, so they can be used for those who have them, and begin testing them. So I think we'll learn more over time, but not everybody has them.

But for those who can in fact pick up a trial use measure, test it, continue to

support their ongoing use and implementation, it is not an expectation that everybody now could take this measure and run with it.

MEMBER ZUN: So are there any vendors of electronic medical records that could use this and give data? Because all I heard was that they're not ready yet. So are we doing the horse before the cart or the cart before the horse?

MR. WALLER: The horse is definitely in front of the cart at this point, and so just -- so Epic, one of the largest providers of electronic medical records, we've already implemented this in our hospital system and we're beta testing it now, and it works pretty clearly.

What we've done, I mean I can -- you want me just to -- we can wait. So I can explain that, how that's done and it's actually working pretty well.

DR. BURSTIN: Just to provide a response to Leslie, because that's a really

good question. I think what we learned is it's also a bit of a chicken-and-egg phenomena. So the vendors need to increasingly know whether the most important data elements they should program into their EHR, since some of that comes from knowing what the most important concepts are.

So some of this is also the push-pull here, quite intentionally making sure there are in fact vendors who will add these data elements to EHRs because that's what we want --

CHAIR PINCUS: That's very helpful, because I think one of the questions I had also was well why does NQF need to do this, then, and as you explained, it really has to do with helping, really, vendors and measure developers to focus, and not really focus on things that were not really worthy of putting effort into. Les?

MEMBER ZUN: As a follow-up to that first comment, so I presume the vendors

have a professional organization. Has it been vetted to them and is that --- I didn't --- is that somewhere in our documents?

DR. BURSTIN: We've extensively worked with the vendors on this approach going forward, and they're very interested in fact in this approach, because it allows it to be more iterative.

I mean we recognize, you know, you want to just throw everything in there, or do you really want to make sure you're putting in what's most important? Even when we did our eMeasure feasibility project, for example, last year we had Epic at the table. We had a couple of other vendors as well, specifically to make sure we're going down a path that logically makes sense for all sides.

CHAIR PINCUS: Jeff.

MEMBER SUSMAN: I'm only going to say if you think about our dsiscussion of the suicidal risk yesterday, I think it's a really good argument for going down this pathway, of

really teasing out a specification related to e-measurement. So is it just the checkbacks? Is it checking down certain behaviors?

How is that reported as we get more sophisticated, can actual language processing, being able to extract data from the written note and so forth, all of which I think would be helpful to have before it comes to this group for more formal evaluation?

CHAIR PINCUS: And that's an important point, that it would come back here, and because of the timing that's expected, is it a routine expectation?

FEMALE PARTICIPANT: Three years.

DR. BURSTIN: We gave the developers up to three years to return. We hope they could do it sooner but, you know, in some spaces truly, for example, some functional status measures that people are trying to test right now, it's going to be a heavy life to find an EHR to do it.

CHAIR PINCUS: Okay, good. So why

don't we get started? Okay. Oh Mady.

MEMBER CHALK: I'm hoping that this approach will be taken with lots of other measures. I really -- I think it's very useful. I think it does push the fields, whatever fields we're in, to pay attention to what's coming down the line, and to work with us. So I really appreciate this.

CHAIR PINCUS: Let me ask one other further clarification, Helen. Measures that are not currently specified as eMeasures, but are already endorsed, do they have to come back? Do they have to come back to be re-endorsed once they're an eMeasure, to go through this process?

DR. BURSTIN: Yes. Essentially the answer is yes. As they come back up for maintenance, we will look at their eMeasure specifications. In the early days of this conversion, there was a lot of PCPI and NCQA, I know this well, this idea of just retooling, taking exactly what the measure was, in

whatever form it was, and just converting to whatever point, you know, to whatever data made sense in an EHR.

I think what we've learned over time is that's kind of really pounding a square peg in a round hole big time, because in fact to develop an eMeasure, you want to take advantage of the eMeasure, and what's the best, or even of what an EHR and other electronic data systems can bring to the table, and starting with the idea of this is what we did in claims and just kind of making it what you do in an EHR doesn't make sense.

So we're hoping we've gotten the, you know, the developers are no longer constrained by saying and it matches what the claims based one says. It's really what makes the most sense in the EHR context.

CHAIR PINCUS: So they would be actually endorsed as two separate measures in principle?

DR. BURSTIN: We're working

through those issues.

CHAIR PINCUS: Okay. So why don't we get started? Oh ---

MEMBER ROBINSON BEALE: Just one point of clarification. Then is there an entity that will take a look at the core data elements, so that there's some standardization around the data elements that go into the electronic health records? Is that the role of the electronic health record trade organization? Is that a role for NQF?

DR. BURSTIN: Good question. So there's actually a fair amount of work being done right now to think that through. There's actually something called, through the National Library of Medicine called the Value Set Authority Center, and the value sets are essentially the way to -- I know, with that look from Andy, it's a funky description.

But essentially value sets are the way to bring together all the codes for something to describe something. So this is

addiction. This is what a visit looks like, things like that. So the measures are -- we're trying to increasingly make sure that those value sets get into the NLM, and there's actually a process for collectively thinking about that.

What does it look like to harmonize those, to make sure they make sense across measures?

MS. HANLEY: Yes, so there's a lot of work ongoing in the standards field nationally, to look at how to ensure that we're representing the data elements in a consistent manner. So we use a data model called the Quality Data Model to categorize our data elements that we're looking for to use in the measure.

As Helen mentioned, we're building the value sets in this national repository, and there's great emphasis to not reinvent new value sets every time you need something for a new measure. So all the measure developers

are working together, you know, using this National Library of Medicine-hosted tool as the space to facilitate some of that work, to really see okay, you know, AMA, we've developed some value sets that NCQA might be able to use in their measure, vice-versa.

So we're looking at, you know, what was the purpose? Why was that value set developed? What's the purpose? Does it meet our needs, or do we actually need something a little bit different here, and in that case we then do develop a different value set.

So the measure development community is definitely looking at that.

MEMBER ROBINSON BEALE: So I assume that that would include the use of things like the tests that we were looking at yesterday, suicide assessments and standardize that type of use of tools, validated tools?

MS. HANLEY: Yes. I think to a certain extent, a lot of that is still going to rely on the evidence base to include those

types of tools and measures. But to the extent that we are using those tools in measures, we are representing them in a consistent way.

DR. BURSTIN: I think there's still more work to do specifically on the tool side though, to your point there Rhonda, that I think is not going to be captured by the Value Set Authority Center. I think that's where you're inputting --- around harmonization and best in class --

CHAIR BRISS: And that the value sets are more about -- or seem to me to be more about general concepts that get used in lots of measures.

So it's the kinds of problems that they're trying to solve with that are -- some of us just looked at sets of diabetes measures, and there were like to decide who had diabetes.

People were doing this in like 17 different ways, and that's the problem, more

the problem that they're trying to solve, than how you define a specification intervention, as in the suicide case.

CHAIR PINCUS: Les.

MEMBER ZUN: So I agree this is the right thing to do, but I'm not sure this is the right way to go about it, and I'll explain that, because I've actually been trying to do this with our vendor. For instance, if a patient says they smoke or they do drugs, they get a referral to substance abuse treatment.

But the vendor keeps telling me they can't do an audit for me. So you know, I can't audit it electronically. I have to manually go through and audit it. So there must -- I know it's crazy, but that's, you know, like order sets. Are the order sets being used, well I have to manually go through every single patient to see if an order set was used for a certain condition.

So what concerns me is rather than

going, working with one vendor or another vendor, there must be an organization, another way to do this through some, whoever sets the standards for electronic medical records or, you know.

Is this the right forum to push the agenda, to get these folks in the same place, so they can help us do the quality measures we want to do. I'm not sure that setting up an eMeasure does that.

CHAIR PINCUS: Mike may have some comment on this.

MEMBER LARDIERI: Yes thanks, and you know, with that, it's really of a vendor problem. Any vendor can do this. If you're looking for them to do it, pick it from your note, no vendor's doing that right now. You need natural language processing to do that.

But if you have a structured data field that the patient says they smoke, that is easily computable to make a referral. So I don't know who your vendor is or how they're

trying to do this, but that's easy. Any vendor can do that. You have to pay them whatever you have to pay for that little piece, but you have to pay everybody anyway.

So even going through this process, you're still going to have to pay the vendor to do it, unless -- until it becomes a national thing that they actually put in place that everybody is doing. You're paying your vendor to do it, right, and they want to do it for you.

So it's not a -- it's really just a vendor reluctance issue versus lack of capability.

DR. BURSTIN: I'll quickly respond to Leslie's issue. So we have been doing this collaboratively with the Office of National Coordinator and very supportive of these efforts. I actually just finished my tenure for the last two years as chair of the Quality Measures Workgroup for the Health Policy Committee.

So this is something very much done with the vendor committee, with the leadership at ONC. I think it's an important -- those are all really important issues, and I guess the question is can quality measurement be kind of a way to help push some of that, and if you can play a role there, I think that's our hope.

But really good thoughts about keeping it all connected.

CHAIR PINCUS: And I think, you know, I mean really the lead for a lot of this stuff is with ONC. Like they are developing, for example, standards for behavioral health EHRs that are not mandatory, but that are sort of recommended and sort of to try to main that standard setting. Like one last comment on this general issue.

MEMBER LARDIERI: Yeah, and then the best place to go to deal with the vendors is through the HINS EHR Vendor Group. It's a specific group that represents almost all of

the vendors in the country. So that's the best place to be.

CHAIR PINCUS: Okay. So let's now talk about the measure under review.

MR. WALLER: All right. So we've talked around it a lot, so let's get down to it. So I'm Corey Waller. I'm a physician. I still see quite a few patients, including a couple on my cell phone last night.

So let's run through what the basis of this is. So this is a composite measure. It has three focus areas. You could call it four, but it's three main focus areas and four fields that would need to be identified.

This looks at screening and brief intervention. It really is a measure dedicated to those two things happening in concert, and it's looking at tobacco use, alcohol use, illicit drug use and prescription drug abuse. In order for us to evaluate whether or not those are happening, we have to

ask those questions, and those questions are not being asked consistently within a primary care space.

So this measure very specifically was developed so that all of those would need to be asked, and then a brief intervention would need to be done, in order for it to be completed as a task.

Ultimately, there's a large amount of research that backs up the utilization of this for tobacco use and the prescribing of medication for tobacco use disorder, as well as a growing amount of literature for the alcohol space, that shows a significant reduction in at risk drinkers, decreasing the amount that they drink.

Following that, we have a little bit of a mixed bag of information concerning the illicit drug utilization, but that seems to depend a couple of main issues, and one of those is how severely ill are the people that you're asking the question to.

We find that many that find no benefit from the brief intervention aspect are seeing patients who are sicker or homeless. I was talking with one of my colleagues this morning, and said instead of brief intervention and referral to treatment, it should be screening, brief intervention, referral to housing.

But at the same time, we do have really good significant data at large scale implementation pathways. The New Mexico study showed a half a million people that they were able to screen and do a brief intervention on, and then also refer to treatment showed a significant improvement.

So they did that in an area which is mostly rural, which is not fully implemented on EMRs, and they were able to show good efficacy and implementation and good outcomes. So what this really focuses on is using a known, standardized screening tool, that would allow for people to choose

whichever tool that is.

So we're not specifying which tobacco screening tool or which alcohol screening tool or which illicit drug screening tool, just one of the validated tools are completed, and then once that is completed, that a brief intervention is done if they are positive on any of those tools.

Once that is done, then that would complete the measure. An example would be if a patient smokes tobacco and has more than, you know, 15 drinks in a week, then they would receive a yes for those two. It would be negative for the illicit drug screen and then they would do a brief intervention that covers both of those topics.

The way that that's captured within an EMR is the screen sits within the EMR, and when that's finished, it shows up as completed, and then the brief intervention is a code. I mean ultimately it's charged out as a code. It was completed and that's how it's

tracked, because those two things are matched.

The way that we've set it up on our EMR for trial, just to answer to Rhonda's question, is very specifically, you can't check the brief intervention unless you've done the screen. So you can't even charge for the brief intervention on that visit unless you've completed the screen to do that.

And for this one, the composite measure, there would be whichever screen it is, screens for all of them. So you would have to finish the entire screen for tobacco, alcohol, illicit drugs and prescription drugs, prior to even being able to code for the brief intervention. Then what that does is it keeps it from being a push-button or a check-box measure.

So that's the approach that was taken by the group that put this together. This was started and pushed by NIDA, as well as SAMHSA. It was brought up on HHS as one of the high, you know, areas of need, as far as

being able to evaluate this.

But ultimately, implementation has been successful in a number of different areas for specifically, the large scale application of this pathway, and what we would be doing is just specifying that all of it needed to be completed so that it wouldn't be chosen, if you only feel comfortable talking about tobacco or only feel comfortable talking about alcohol, that we would miss one of the bigger issues.

MS. HANLEY: I would just like to add that the alcohol and tobacco components of this composite are existing NQF-endorsed measures already in use.

CHAIR PINCUS: So Peter, you're the lead reviewer.

CHAIR BRISS: So as you've heard briefly, this is -- all four of the proposed components of this composite are unarguable big sort of public health burden issues. As you've heard, there's a performance gap in the

sense that all of the issues are under-ascertained. Even the ones that are currently better supported, tobacco and alcohol, are still under-ascertained.

The evidence to support the measure focus. This is actually a good measure that tests the approval of testing, for the evidence to support the measure focus does have some challenges, I would say.

So tobacco and alcohol have been well-studied. It's clear that screening and brief intervention improves outcomes. They're both recommended by the Preventive Services Task Force. The other drug measures, whether you consider them one or two, are mostly untested. They're mostly unrecommended by the task force and others.

The weight of the evidence, as nearly as I can tell, is that screening and brief intervention actually hasn't worked. It's not just on the study. It actually hasn't worked, including two recent good

studies in JAMA.

And so the -- so there are legitimate questions, I would say, about whether this would have a net positive or net negative public health impact. You could make an argument from where we are today that this would move forward.

All four issues, you could make an argument from where we are today that the additional burden and confusion introduced by adding other drugs could actually have a net negative effect on tobacco and alcohol screening.

The other thing that I would say, that testing the measure doesn't actually -- it might help you assess what the net effect is on performance of screening, but it won't help you assess what the main question is, which is can you find a screening and brief intervention that actually works for other drugs, right?

So testing the measure won't

actually answer what seems to me to be the core question, and I guess I'll stop there.

CHAIR PINCUS: David.

MEMBER PATING: No. I just want to add that I do think this is the perfect kind of subject to do a trial measure on, because I think we really just need to answer the question, and the question is can we do screening and brief intervention for the drug component, and does it make sense to do it in the context?

Because we know that there's just a lot of overlap and comorbidity between alcohol and drugs, and so I think you're really testing that component.

Regarding the JAMA article that just came out, that was a 40 percent homeless population. So I really don't know if that should weigh negatively in terms of the ability to intervene in screening for drugs, as well as having interventions. It's very, very -- was it an insured population? I'm not

even sure they're a public sector population or a Medicaid population.

So that again, with regards to this, I think these are the open questions that make this, you know, ripe for making this a trial measure.

CHAIR PINCUS: So other comments on this issue with regard to evidence around the measure focus and importance?

MR. WALLER: I just wondered if it would be okay if I spoke to the specific articles. So I called Rich Sates, so that we could figure out. I needed to understand this as well, because I came at this measure late. I wasn't the developer. I'm helping shepherd this through from a clinical perspective, and so I needed to know this.

So a couple of things about the information that's come out, is the sites that were utilized in that have already done three other trials with SBIRT. So their control group is going to probably deliver SBIRT at a

higher rate than most people who are doing it as the experimental, meaning that their control group is already getting the intervention on a regular basis, and it just wasn't a track. So that was an interesting thing, is that -- so they were sicker, but at the same time, the control group was a little better.

So I just -- this has been an interesting thing to track, and there's good data on large-scale studies and on the smaller-scale studies. They seem to not have the efficacy. I think there's a good question to answer, whether brief intervention is a good use of time for this, and that's where I think that this measure gets to the crux of it.

CO-CHAIR PINCUS: So one thing actually, just before we get to people's comments, just to clarify, and it may be useful as, you know, in thinking about this process. What exactly will be tested? What

are the sort of key elements of the testing, and how will -- because in a sense, we're approving this for testing, and it seems to me that we ought to understand what it's being tested for, and in a sense, what hypotheses are being projected within the testing process, and under what circumstances would you then sort of come to a conclusion that you need to modify this?

MR. WALLER: So I'll speak first, and then they're going to have some extra comments on this. But the screening piece is not the piece being directly tested. We know that screening and referral, the treatment for illicit drugs has a significantly positive impact.

What's being tested with this is, one, can this be implemented in large scale and consistently evaluated through an eMeasure. So there's that component of it as truly can we see this happening and watch the outcomes from that, and the outcomes would be

are we finding a decrease in utilization on retesting for these patients.

CO-CHAIR PINCUS: So that's going to be part of the testing?

MR. WALLER: Well, that's going to be the determination. So there's the testing piece that tells us whether or not the measure is implementable and usable and consistently valid within the system, and then the second piece is to determine whether or not it has efficacy.

That efficacy is whether or not the brief intervention makes a difference for a decrease in the prescribing for nicotine, as I stated earlier, which I don't think will happen. I think we have some good evidence to show that that's not happening.

But the other thing will be does it decrease the utilization of the substance that was screened for and then intervened upon by a brief counseling session, and determining whether or not that had a significant impact

on the utilization.

Or do you just not brief intervene once you've screened? We have to answer that question. Do we add another thing to a primary care doctor's docket if it's not working, and if it's not working, we need to have that understanding that those patients should be referred for treatment after a brief discussion. So the brief discussion and intervention --

CO-CHAIR PINCUS: So wait a second. So you are in fact going to be testing the efficacy? That's the plan?

MR. WALLER: Yes. Yes, sir.

CO-CHAIR PINCUS: Okay. So why don't we get started? Why don't we sort of work our way up here and then go around there? Okay. So let me see. Who's at the end there?

MEMBER MAZON JEFFERS: So I'm really glad that you said that this measure is about testing the brief intervention component, because actually one of the things

I found interesting about it is that the specifications of what a brief intervention is, it's incredibly vague, right. So it's a doctor having a chat for five minutes, and there are evidence-based brief intervention strategies that we know to be effective.

I also think, I'm having trouble re-finding the measure specifications. But for tobacco, I think you can either have a counseling session or do pharmacotherapy, and we know that the intervention that is most effective, based on the evidence, is the combination of the pharmacotherapy with the counseling session.

So I'm just wondering if the point of testing the measure is to better understand the impact of the brief intervention. I'm wondering why the specifications around the brief intervention aren't more explicit and don't require more adherence to what we know to be the evidence for that piece to be effective?

CO-CHAIR PINCUS: I'm wondering whether given the number of comments, that we should go through the comments first and then have you respond. So next, Tami.

MEMBER MARK: Just a point of clarification. So it seems like the validity issue, it's most of concerning to the illicit drug abuse screening brief interventions. So are these -- it's a composite measure, but each of those separate types of screening and intervention are going to be tested separately?

CO-CHAIR PINCUS: Now let's keep track of these, okay.

MEMBER SUSMAN: So two questions, one very nitty-gritty and one much more conceptual. The nitty-gritty one is I was trying to understand the specification for misuse of prescription drugs, and is there an evidence base around that, and this particularly in an electronic record could be difficult to tease out consistently.

And then the more conceptual one is maybe I'm not understanding how this measure will be reported, that as I'm reading it, it's going to include those who are screened and are negative, and those who are positive and have an appropriate follow-up. But if you have different rates of substance use in your underlying population, it's going to be very hard, I would think, to control for that different prevalence.

The hard part of those who screen positive and get a brief intervention whereas if you're actually toting up everybody, and there's only a very small number of drug users in one population, a great number in another population, it's going to be carrying apples and oranges.

That's more a matter that could be easily cleaned up with the actual specification of reporting. But at least as I read it now, it sounds like that would be sort of confabulated within it.

CO-CHAIR PINCUS: So just to add to that, I mean the example of that is in a study that we did, evaluating the quality of mental health services in the VA, in which we compared that to the databases. This is the administrative database, comparing that to a private sector database.

The VA actually did somewhat better, in some ways considerably better on all measures, except for those initiation engagement measures and substance abuse.

The reason was is that we think that the VA, in the population we looked at, had a 23 percent prevalence of substance abuse, and the private sector had one percent, and the VA was doing 100 percent screening, and they were identifying people with much less motivation, and likely they were less likely to follow up.

So there it is. So I don't want to extend it. It's just in responding to that sort of thing, just how you deal with that

problem.

MEMBER SUSMAN: Thank you.

CO-CHAIR PINCUS: Okay, Connie.

MEMBER HORGAN: Thank you, Harold, for asking your question about what exactly can be involved in testing, and this question relates to how creative can one be with testing.

My understanding is that the efficacy on the drug measure really relates to it being a universal screen, and is there any opportunity for creativity for defining the -- who is tested by perhaps high-risk populations or patients who have screened positive on alcohol and tobacco?

Is that kosher in composite measures, to basically change what you're -- the base across the three measures? Because I think that one problem with the drug measure has been related to whether it's universal or not, and is that something that can be considered, or is it even kosher in a

composite measure to do something like that?

CO-CHAIR PINCUS: And then -- Mike. And then we'll give you a chance to respond to the questions thus far. Right now, we'll go halfway and then let them respond, and go on. Because people won't be able to keep it in their head.

MEMBER TRANGLE: Okay. I'm coming from a practical perspective probably more than theoretical. We've been doing this for a number of years, you know, starting way back when with ERs and trauma units in hospitals, and expanding it to primary care clinics and then health plan case managers telephonically doing kind of SBIRT for folks with alcohol and opiates and soon to be benzos.

And as we've kind of worked with this, several thoughts occur to me. One is the fact that you're trying to get data from an EHR/EMR is really much better than trying to do it any other way. And to some extent, finesse, the issue that's plagued us forever

about who counts?

Does it have to be a billable provider if it's claims-based, you know? When most of the evidence is you can take lower level people that don't need to really be billable providers to do this, as long as they're supervised and trained. So you finesse this huge issue that's beleaguered, I think, a lot of people trying to work on this.

Two, in our state, as in a number of states, there are a lot of sort of coordinated initiatives going on, to sort of stem the tide of prescription drug abuse.

We're finding that we're getting much better. I mean clearly a patient, if a prescription drug becomes hard to get or becomes too expensive, they'll just buy heroin cheaper on the street. So I know that there's a lot of flow back and forth.

But we're beginning to get a handle on what's going on through our EMRs with prescription drug abuse that's really

pretty accurate, even though we don't totally get the sliding over to the guy on the street stuff. And we're having reports that we're starting to routinely see about, you know, whether they're sticking to our protocols and whether you can capture it automatically.

I would encourage you to think about, is this a bite-sized, actually very feasible, doable piece that can show a big improvement, even though it's not all illicit drugs lumped with it, you know, because I think there's room to grow there, and maybe I'm also going against the room as a caution here, and well, I guess those are my two main thoughts.

CO-CHAIR PINCUS: So if you could respond to those comments thus far.

MR. WALLER: All right. So evidence on the type of brief intervention. So specifying the type of intervention. When I went through the list of who sits at this table, I saw a lot of psychiatric colleagues.

So the reality is the brief intervention, while there are standardized approaches, they sit under an umbrella of, you know, cognitive behavioral therapy, motivational interviewing and even mindfulness aspects of approach.

So that becomes a delicate balance between how well you know the patient and anything, because you may know the patient where the brief intervention is. You know your wife wouldn't be very happy with this, and that may be more effective than following a full-on CBT.

I've had this conversation with now 85 primary care doctors. We have -- I work for a group that may have 400 primary care doctors in this medical group, and in having this conversation with them, they're going to brief intervene based on how they feel is best for the patient.

We all know that that's still, you know, throwing a handful of rocks at a target. So the reality is, is that I don't think that

we can truly move that needle without at least starting with the -- do the brief intervention, and then those that feel very uncomfortable will go to the education of how to do that, because we supply that education for what are the standard models.

But that's the one piece that's going to be difficult, is what brief intervention made the difference. So if you have one provider that falls out with more positives than another one, we're going to be able to figure out, because it is an eMeasure, because we can actually now look at population health statistics over a period of time to determine which providers do it better as compared to those that do not.

So I think early on, trying to specify, to over-specify the brief intervention may be actually a mistake, given the significant heterogeneity that exists within practice styles and approaches, and the fact that the brief intervention data is also

plagued by cultural inaccuracies, which populations it works best in and which populations it doesn't.

So I think it also is one of those things that while tested, is only tested for the very specific populations that don't exist and from the peer world that we're trying to do that.

Will they be tested separately? So it depends on how the EMRs bill them, to be honest. I mean if the EMR billed it as a package, and so if they billed it as a package, then it's going to be difficult to test it separately, because what you're going to get is a yes these were done as the measure comes out. If it's built --

CO-CHAIR PINCUS: Well, can't you specify about how -- I mean I guess, and Helen may want to explain to us. When we're approving a measure for testing, is it being tested along a particular protocol or it's totally up in the air what that protocol is?

Because it seems to me that -- I can't speak for the whole committee -- but what people move toward is if you want to test it, you want to test it both as a, you know, either as a composite or as separate ones, so that you modify it, given the fact that the weakest link is the drug abuse piece.

DR. BURSTIN: It's a good question, Harold, and I think some of this is I think when we were thinking through trial measure use, we didn't really think some of the first measures out the box would be composites.

It surprised us, and in fact this is -- we've done this is in musculoskeletal and they are all pretty simple measures. So this is a little bit of a challenge. Our current composite measure evaluation framework really requires the analysis much more sort of be at the level of a composite.

Now in this case, being a trial measure, we probably need to know a little bit

more about the validity in an individual -- at least the ones that aren't already endorsed. We could talk with PCPI on that.

MR. WALLER: In Epic, we're building this as each one individually and a composite, so it can be broken out. So I know it can be done, and they're connecting that still to the charge code of a brief intervention so that they can track it. That's any brief intervention within a defined office set.

So if it gets referred to a social worker for the brief intervention as compared to the provider, that will still count and it will connect, because they've connected that to the EMR.

DR. BURSTIN: But you know, one of the requirements is that composites can be unpacked, for the sake of quality improvement. I think they're going to have to be built individually.

MS. HANLEY: And also because the

individual data elements for each aspect of each of the measures will be collected separately, we will have that capability --

CO-CHAIR PINCUS: So we can -- so in principle, then, we can endorse this with an assumption that they will be able to be unpacked?

DR. BURSTIN: Yes.

CO-CHAIR PINCUS: Okay.

MR. WALLER: The next one was the evidence base for testing for misuse of prescription drugs. So we're at the point now where we can almost tell this without talking to the patient, I mean like you were stating. I mean so when you look into, you connect the electronic medical record with a health information exchange with a prescription drug monitoring program.

With the PDMPs, I mean you can pull a prescription drug monitoring program report, in all except for Missouri. And so in every state, you can pull that out and look at

this, to determine whether or not they're receiving prescriptions from other places.

You know internally whether or not they're calling early for prescription refills and things of that nature. So it's actually one of the easier things to tease out in a single practice nowadays. The question pathway is not as well developed as that for the misuse of illicit drugs, I mean, because that's what's been focused on for so long.

Because up until really the mid- to late 90's, we didn't write so many prescription drugs. That became the -- right. So in the end, it's the single-question effect has been shown to be really the easiest. In the last three months, have you taken your prescribed medication for something not indicated on the prescription, or for a reason other than it's prescribed?

That single question has been shown to be just as valid as any long-term input. So that's validated against known

mechanisms, but that single question seems to be just as valid. So that's why I think that this one has the ability to tease that even better.

We're testing this measure and, you know, as I say right now, in a couple of the pilot locations, we've already seen exactly what was seen earlier, was talked about earlier, in those pilot locations, a per person decrease in the opioids prescribed for chronic pain.

That is directly correlated with the percentage of the patients you screen positive for either illicit drug use or misuse of prescriptions, or alcohol or tobacco. So even the screening for those two have significantly impacted that with the education, that having a use disorder with any one of those increases your risk of having that.

So the screening piece has had really nice secondary effects already, and I

think a number of other people have shown that the screening by itself changes behavior for other things that negatively impact health, like prescribing an opioid to an opioid addict.

CO-CHAIR PINCUS: And Jeff's question?

MR. WALLER: That was the how reported prevalence?

CO-CHAIR PINCUS: Yes, the reported prevalence.

MR. WALLER: Yes. So if somebody says that they only have one percent of their patients that have a problem, they're not screening well. I mean I think that what we're -- what we're going to be able to do with this is also benchmark.

So what we haven't been able to do is actually benchmark the data set across hospital systems and providers, so that if you see in a population health statistic that you're two standard deviations from the mean,

you can guess that you're not doing it well.

And it's not a matter of you have the best patients in the world. It's very much a matter of you're not asking the right questions in the right way at the right time. So I think that this is --

CO-CHAIR PINCUS: Will you be able to report the percentage of people who are screened?

MR. WALLER: Yes, absolutely. That's really important with this, because we need to determine --

CO-CHAIR PINCUS: The percentage of people who screen positive?

MR. WALLER: Yes, of those asked, of those screened for tobacco.

CO-CHAIR PINCUS: And that's part of the report that comes out of this?

MR. WALLER: Yes, and it's really one of the most important pieces, quite honestly, because we haven't even been able to benchmark this, this effect throughout, which

is why we sit at this abysmal ten percent seeking treatment number.

CO-CHAIR PINCUS: And so that, you're saying, in a sense is a balancing measure, as well as a performance measure?

MR. WALLER: Right, and so putting this out for testing in this way, I think, is really important, because it allows us to at least level set, you know, hospital systems and payors even to level set if the dollar they're spending is giving them the value for what they hope it will be, you know, in evaluating these patients for these very high risk.

The two biggest preventable causes for emergency department visitation is alcohol and illicit drug use. So this is a really important piece that we're going to be able to find with this measure.

MEMBER SUSMAN: But certainly it's possible, in fact probable, that certain populations have an increased risk of

underlying, for argument's sake, prescription drug abuse, or drug use overall. So if I'm down at the homeless clinic seeing my patients, I think it's probably reasonable to expect the rates of screened positives, the overall true prevalence is a lot higher than some other population. Not that I might not be missing some in either area but --

MR. WALLER: And I think we'll be able to find that and we'll be able to define that based on payor now, that the EMRs track that as, you know, it's beyond the specific data point.

So everything that's entered into a specific field can hold. So we're going to be able to really define commercial versus public, you know, and private pay versus male versus female versus age range versus white non-Hispanic, Hispanic, Asian, African-American.

We'll be able to pull this data. All that stuff we can now start to cohort.

But at the same time, while those things are expected, I found that -- because one of the places that we piloted, that we're piloting is one of our good, nice places, and we're finding 16 percent positive rates in a working group with all commercial, with all of this and the primary care doctor calls me freaking out like twice a week now.

What do I do? My 72 year-old patient is smoking hooch on the side. I mean so I mean -- I mean but that's -- but these are things --

CO-CHAIR PINCUS: Tell her to smoke more.

(Laughter.)

MR. WALLER: But the reality is is that those are things -- yes, the highly technical term as a neuromolecular biologist.

CO-CHAIR PINCUS: Okay. So let's go -- let's now hear from Mike, Rhonda, Peter. I have a comment and then Raquel has further comment.

MEMBER LARDIERI: Great, thanks. I guess my question is, is the brief -- does the brief intervention have to happen on the same day as the screen? Because if you're dropping it to a code to bill for SBIRT, you have to do 15 minutes, and that could be cumulative with all the staff, including the M.D.

But if that's what you're using, most patients, the first time you talk to them, they don't want to deal with it. So you deal with it next time, and that's where the whole motivational interviewing stuff comes in, and maybe three times down the road before you're actually able to get your 15 minutes of intervention. So I'm wondering how that plays out.

MR. WALLER: I can answer that quickly. It does not have to be on the same day, just within the specified 24-month period.

CO-CHAIR PINCUS: Okay, Rhonda.

MEMBER ROBINSON BEALE: This is a very complex issue, to say the least. I just have a couple of comments.

On the drug use issue, National Quality Forum several years ago sponsored, I believe it was with NIDA, a national standards work group, and out of that there was -- I was co-chairing -- and there was a lot of argument around the whole issue of screening the general public around drug use.

I think from that, there was a recommendation for -- and it was heated, let me tell you. It was not a fun time. At that time, the recommendation was for high risk populations, and even there was argument in terms of defining what the high risk populations were.

Certainly those women who are pregnant was high on the list. Adolescents were high on the list. But then the issue of other types of populations really got into kind of issues of profiling populations and it

got kind of ugly. So that is when it became unclear.

I'm wondering whether or not with this measure, since there is that controversy. There's a lot of issues around this, whether or not what Connie was kind of suggesting, in terms of if this, then that, or if you have an identified high risk population, pregnant women, women who are pregnant and adolescents, there's enough data to suggest that there is high issues in terms of mortality and morbidity, particularly with women who are pregnant and with adolescents, the high prevalence.

If that might make an alignment at least with the guidelines, and therefore may be in a way more acceptable. Around the brief intervention, I just want to reiterate what you said around brief intervention, because it is -- falls in an area of psychosocial interventions.

You're absolutely right. There is

so much in the components of that, that makes it kind of -- it may appear to the general public as being kind of squishy when it really isn't. There are core elements within that that are still yet to be defined, and there is a committee that's kind of trying to work on defining that.

One of the issues is that the components of that are not necessarily teased out. Like therapeutic alliance is a very core part of this, because if you have a good alliance with the patient, you can hopefully influence.

But with that being said, I just want to support the issue that it has to remain somewhat vague like that, because it's not very well defined in the field.

CO-CHAIR PINCUS: David.

MEMBER PATING: As I'm actually listening to the discussion, I'm getting more excited about this measure and the possibility of answering some really important questions.

So my system has done alcohol screening now for three million people as a routine across all primary care services.

We continually do alcohol screening. We didn't go with drug screening, because we just didn't know the impact of it. So alcohol screening, we rationalize it, saying alcohol screening picks up 75 to 80 percent of all comorbid drug use.

The question that I would be really interested in studying would be does alcohol alone, is that adequate? Is alcohol screening alone? Does alcohol plus drug screening add anything, or just drug screening alone pick up something that's more significant?

So I really like this combined as a measure, because I think you can't tease them out because of that. But that's the research question, is whether this adds extra value or not, and I'm just really excited to be able to dig into this data set, including

pulling out the pregnant populations, the young adults and seeing if it's high risk, if it should get a triple whammy as well, some other thing, you know.

So I just -- you don't need to comment there, but if you can later, just you know, the ability to segment these things just makes me very excited.

CO-CHAIR PINCUS: Okay, David.

MEMBER EINZIG: So I also think it's a very important measure. But my question has to do with I'm trying to think about this from a patient perspective.

So I'm a patient and I've been to my clinic twice in the past year, and I see, you know, my primary docs. I'm healthy, and I just see whichever doc, whichever provider is in the clinic.

So I'm going there for a sinus infection, sinusitis. So this screen is going to have to be applied to me. So I'm just trying to wrap my head around that. It seems

like a pop quiz, where if I'm going in for a preventive visit, I expect it.

MS. HANLEY: So for the eligibility --

CO-CHAIR PINCUS: Can we go to other questions. Let's just kind of get through and then respond, just to keep it efficient. Larry.

MEMBER MILLER: Yes, I think this is going to be a very important measure. The question I have, and it's probably a very simple one, I understand that the intervention is met by a code, a CPT code or whatever that's done.

How would the screenings document that check that's been done is actually the results of the screening, so that you can look and see how those are done, and for different populations. How is that sort of handled?

CO-CHAIR PINCUS: So why don't we have you respond now after that, and then I have just one small thing and maybe Peter, you

could sort of summarize.

MS. HANLEY: So the -- regarding the intervention, a brief counseling, a brief intervention here is allowed. So it does not need to meet the threshold for billing for that service. So it can be a minimum of 5 to 15 minutes. The other -- can you repeat the second part of your question?

MEMBER MILLER: I'll try. With that follow-up, so how has that even been documented if it's not a billing? But that wasn't the question. The question was about the screening, how is it done.

MS. HANLEY: The screening, thank you. So those -- that information is captured in the electronic health record, with the result of the screening. So if you're a tobacco non-user, if you don't use drugs at all, those answers are captured as part of the screening, and that's required as part of -- to be able to report it from the EHR.

MR. WALLER: And that's the way

that we're building it right now in Epic, and Cerner is actually paralleling this at the same time, because we have both of those systems in our hospital, which is awesome.

But the reality is we're able to say on the soap, you know, if we're going to do an evaluation, then what we can do is say Question No. 2 is answered positive by more of these patients. So we can start to even validate some of those subsets of questions.

It's really going to be helpful for us, because we're making it -- so those that screen positive are actually trying to validate the 11 subsets of the DSM-5 for SUD on those independent factors, and those are being mirrored as well.

So yes. I mean we're going to be able to cohort this data to death. I mean that's the point of it. I mean we want to flog it at the end of the day. I mean it's going to evaluate everybody from 18 and up, and if you come in for your sinusitis, why do

you have that? Are you smoker?

So these are -- it becomes very -- I mean those are the questions that become realistic if you're, you know, the high risk. If you smoke and you have sinusitis, the chances of antibiotics helping decrease by 60 percent. So it becomes highly relevant, even on a very basic issue.

Chronic back pain absolutely does not get better in patients who smoke, you know. Chronic shoulder pain or bursitis, because you get drunk on Saturday nights and you sleep like this the whole time. Those are very specific questions that can -- you can elucidate cause and effect by asking these questions.

But you're responsible for the vast majority of non, you know, emergent visits that come in if you have a respiratory illness --

CO-CHAIR PINCUS: Could we move to, you know, sort of respond to Rhonda and

the two Davids?

MR. WALLER: Yes. So with Rhonda's question, so there's two pieces. One, I see all the pregnant patients in our county who are on controlled substances. So I agree that without question, that's a high risk population.

But what percentage of patients are currently screened in an OB/GYN office? Less than five percent nationally, because there is no requirement to push that out.

So less than five percent of moms are screened for alcohol or drugs. They're screened for smoking, because that's just a part of what's already on the EMR. But in general, less than five percent even do the screen, even on the highest risk population.

So that is definitely concerning to me, and I echo that. I've also been in some of those rooms with these conversations that were heated about -- most of that has to do with a fear of the perception of the

disease, and a worry about creating a those-patients approach when you cohort those, and whether or not this would be utilized as a negative for the patient rather than a positive for treatment or intervention.

This is -- since I've been involved in addiction medicine been Topic No. 1, part of the problem is that this disease is treated in a closet, and getting it outside of that and allowing it to be aired as no different than diabetes or hypertension nor urinary retention.

It doesn't matter. It's the same thing. We have treatments, we have approaches, we have evaluations. But it starts with screening and asking the appropriate questions that are validated to give us an answer that we know it needs intervention.

So I think that this is an easy Step 1 in making the disease and the risk of the disease within a population visible, and

I think that if we're unable to take Step 1, Steps 2, 3 and 4 will never happen. So my concern is that the risk to our population by not asking these questions is made evident by the fact that we have more people die of overdose than car accidents in three-quarters of the states.

So the reality is not asking that question and then writing a prescription for an opioid or benzodiazepine or a combination of those two that causes all of this problem is, you know, quite honestly it escapes me why this hasn't already been asked for a decade.

I know the answers to that, but I think that this puts us in a position, and in really a wonderful timing situation, to be able to drop this into an electronic stream, so that we can pull population health data, which is going to allow us to understand on a large scale the impact, rather than, well, when I talked to my ten patients, I had a really good effect, or it didn't seem to work

for me, because that -- quite honestly, that anecdotal stuff doesn't really matters. It matters to the physician who gets the data, that says you need to figure out how to do this better or good job, continue to do that and maybe you can help your colleague.

But I think the reality is, is looking at the population health data also takes out the individual bias that seems to be worried about so much, because as this gets dumped into an EMR stream, you're going to be able to look at data sets and not people, and that becomes helpful for disseminating.

CO-CHAIR PINCUS: You need to shorten a response a little bit, and get to the two Davids' questions.

MR. WALLER: Sorry.

MEMBER PATING: Actually, I'm fine.

CO-CHAIR PINCUS: You're fine? Okay.

FEMALE PARTICIPANT: Well, if we

could just explain the denominator population. Oh sorry.

MS. HANLEY: So for preventing -- if a patient's coming in for their annual wellness visit, that visit alone will count for inclusion in the measure. Otherwise, we require two other types of visits. So again, if you don't address it at that one visit, you do have the opportunity of another one.

(Off mic discussion.)

MEMBER EINZIG: Just the complexity of if a patient sees a different provider at the same clinic, not a primary care provider per se but a different provider, it's --

MS. HANLEY: Is your concern that they would be included or not be included?

MEMBER EINZIG: Well, I'm just trying to think. So the process of patients coming in for their sinus infection or what-not. They're seeing the doc de jour. You know, they're not necessarily going to be

looking to see if the screen was done in the past 24 months. They're just there for a sinus infection, and they have no other history with this patient. So just real world scenarios.

MS. HANLEY: Yes. So it would -- I mean because this is specified at the individual clinician level, it would require two visits with the same provider.

CO-CHAIR PINCUS: And Raquel had sort of a question, just coming back in terms of your response.

MEMBER MAZON JEFFERS: Okay. So I just want to also say that I do think that this is an incredibly important measure, and learning about the prevalence in primary care settings of substance use and bringing it, as you said, out of the closet is really significant.

I understand that the brief intervention today is more of an art than a science. But I'm wondering if you can't use

this measure in a way to help bring more of the science into the brief intervention, and try and understand a little bit more about what is working in terms of a brief intervention, because the way the measure is constructed now, we're not -- you're not even going to be gathering data on what in fact the brief intervention entailed.

So it might be helpful to at least gather information on what in fact was done during that brief intervention, so that we could begin to accumulate information about what would be effective, what would be more effective, understanding that this today, we're working in more of an artistic world than a scientific world around the brief intervention.

CO-CHAIR PINCUS: Thank you. Peter, you want to summarize? I had a couple of comments, but I could do that at the end, because this actually doesn't require a response.

CO-CHAIR BRISS: No. Everybody's excited about the importance of the issue. I do want to make some comments. So one of them is that I think that the verbal description of the evidence is for the drug interventions doesn't match what's in the printed materials, and frankly what's in the various intervention studies.

So I think I'd encourage the developer to if you have data that isn't written down, that makes you feel more positive about these interventions, I'd encourage you to marshal it.

I do think this is going to be -- as a primary care doc, I'm worried about professional burden of this measure and on an intervention that's mostly been untested.

And then the second thing that I would say is that there's clearly a lot of hope around the table about things that might be tested in measure testing. I'm skeptical, frankly. So think about what actually gets

written in the EHRs.

You're not talking about a cohort setting. You're talking about real life clinical information. If you're lucky, you're going to have somebody say I addressed this issue, right, or have a checkbox that says I addressed this issue.

You will not get information about quality of counseling. You will not get -- you may get information about whether processes were done, but you will not get information about whether people's outcomes were better.

And so we shouldn't over-emphasize the kind of -- some of the big questions about these are fundamental efficacy questions, and you will not get that information from EHR data, any kind of EHR data in any setting.

CO-CHAIR PINCUS: So I think we're ready to vote. Oh Mady, last comment.

MEMBER CHALK: I want to point out to Peter that the same applies to the measures

we discussed yesterday, follow-up for people with SMI on all of those measures and all of those exams tells you nothing about the counseling or the follow-up, and we didn't have big issues about that. So I don't want this particular composite measure to be held to a standard that wasn't used yesterday.

MEMBER SUSMAN: This is probably mostly for Helen. If I heard this discussion during one of our usual conversations, I'd say well, this isn't ready for anything. I mean we got all kinds of questions.

It seems more like a discussion with a very smart research associate, who's got, you know, lots of different ideas and projects they're going to do, all of which I find very exciting, and I don't mean to demean it.

But as far as sort of even close to prime time, I'm feeling really uncomfortable. So I'm not quite sure what NQF's intent for this group is at this point

in time before we vote.

DR. BURSTIN: That's a good question. You know, I think in our expectation, we came up with the concept of doing trial use was the idea that, you know, these weren't ready for testing, but they were ready for everything else.

So I do think you need to have a level of comfort about evidence that is equivalent to any other measure you would look at. So it's not as if you're going to have evidence re-examined as part of this trial. I mean I just -- just to be honest about it, it's not as if that's going to emerge.

We did hear a fair amount about what they could do in terms of the rates of the individual components that I think would give you a great deal more comfort about validity, you know, as much as we'd like to try to tease those apart, are so heavily related to each other.

So I think you need to decide is

there enough that these measures could move forward for trial use or, you know, is there still enough discomfort about the level of evidence of the four components? That's really why we have a standing committee. It's not for us to decide.

But again, keep in mind -- the last thing. Keep in mind that even if they go through this whole process, they are not endorsed. They're only approved for trial use.

So there is an expectation that these are sort of out there on a trial basis, with not an expectation that somebody could turn around and say these are endorsed measures, and let's use them to pay providers, et cetera. You would expect that other work to happen first.

MEMBER SUSMAN: Then could I just ask the measure developer, from your perspective, why do all -- it sounds like you're going to do this stuff pretty much

anyway. I mean does having the stamp of being an eMeasure for whatever this means, I'm sorry, the lingo, forget it.

Does that have enough impact for you as the developer that it really is important?

MS. HANLEY: You know, I think that at the national level, there is huge emphasis on measures, outcome measures, composite measures, patient-reported outcome measures. We're working very hard to try to advance the measurement field towards that, and this is one stop on that journey.

DR. BURSTIN: And just from a national perspective, I'd also add, I'm glad you made that point. There's also just a huge interest in moving eMeasures out. So I think some of this is also just trying to put it out there, see what we learn, rather than waiting for the perfect, which could take a while in the eMeasurement space.

CO-CHAIR PINCUS: So Helen, I'm a

little bit concerned that your last statement, you know, what you just said is in total contradiction of what you said before, well, you should apply exactly the same level of expectation, level. Because the reality is that, you know, there are more gaps in the evidence base for this measure than there are for the other measures.

Part of the thing that's accounting for some of the enthusiasm about this measure is that we're going to get more evidence to be able to fill those gaps. So that if -- so in a way, particularly for this criterion, you know, it's -- you know so just talking in my own head, sort of walking into this, sort of how I was going to vote.

I was walking in here thinking like oh, there really isn't evidence for the drug abuse piece, and I was, you know, sort of kind of negative about that.

But hearing that this is actually going to get evidence that's going to inform

us in a better way has made me more enthusiastic and more likely to vote in favor of it. But now you're telling me that I shouldn't do that.

DR. BURSTIN: Didn't mean to -- yes. I didn't mean to confuse anyone. Let me just try it one more time. At least our initial conceptualization of this, and it's interesting. When I presented on some of this at Academy of Health, I got pushback on exactly the issue.

Is it really just that the testing reliability and validity will emerge from testing, or in this world of electronic data, does evidence actually emerge from the work of putting it out there? I think that's a real question, is you know, will there be real-time evidence collection as a result of having this out there in trial use that would inform you?

Frankly, that's uncharted territory. I mean again, it's trial use. So I think there's a bit more of an expectation

that it doesn't have to be quite as clear, and also at the end of the day, you're not saying this measure is endorsed. It doesn't meet the same standard.

So I'm not saying -- at least in terms of evidence, you may want to, for example, consider the fact that you do at least have an option of insufficient evidence with exception, which is specifically something where you think currently the benefits to patients could outweigh the harms, but you're not completely and totally convinced that it is, you know, right at this very moment something you could move on.

I'm not encouraging that. It's supposed to be an exception, but I just --

CO-CHAIR PINCUS: Well, what does that mean with exception?

DR. BURSTIN: It means with exception, literally that you would say --

(Laughter.)

CO-CHAIR BRISS: It means we want

to -- we believe that it's a good thing to do and we want to recommend it without evidence.

CO-CHAIR PINCUS: Right. So that counts --

DR. BURSTIN: That counts as a yes.

CO-CHAIR PINCUS: That counts a yes, okay. I just want to be sure, because it's listed as a 5, but so it counts as a 1 or a 2.

MEMBER TRANGLE: Why don't you just allow us to review the evidence when it comes back later, instead of saying review it now and never look at it again?

DR. BURSTIN: Because in general, most people will come back with only testing results that will have no impact on your --

MEMBER TRANGLE: I know, but can't we make an exception that way this time, with exception? That's how we define exception in this space.

DR. BURSTIN: I think that's --

that's essentially what this is, it's insufficient evidence. Do you believe that putting this measure out there will have significantly more potential benefits than harms?

CO-CHAIR PINCUS: Out there for testing or out there for --

DR. BURSTIN: Out there for testing. This is only for testing.

CO-CHAIR PINCUS: Okay, oh. Well right.

(Simultaneous speaking.)

DR. BURSTIN: When it comes back -- the question is when it comes back --

CO-CHAIR PINCUS: Let's, let's -- okay. So Connie has a question and --

MEMBER HORGAN: This is a follow-up to Harold, your question, and the issue is impact on unpackability. One concern with composite measures is you get a single score, and how actionable are the components. How unpackable is this measure going to be?

DR. BURSTIN: (off mic)

MR. WALLER: It is unpackable because they each independently have to happen. So we need to show that each of the independent factors happened, and then the brief intervention was done for any positives.

MEMBER HORGAN: Right. How usable will it be by a health plan to use it in that way, and why not have paired measures versus a composite measure, because you know, the ease of unpacking. I mean you could have a locked suitcase that's very difficult to unpack.

CO-CHAIR PINCUS: So let me see. Let me sort of just try to clarify, so that we can move ahead, because time is marching on and people have planes. So it sounds like if we -- that if we move ahead with sort of recognition of this measure for testing, that it could come back to us for endorsement either as separate measures or as a composite measure?

MR. WALLER: I think that yes, ultimately. Because we're going to be able to determine the validity of the approach for each of those independently, you know, as far as --

CO-CHAIR PINCUS: Just a yes/no.

MR. WALLER: Yes.

CO-CHAIR PINCUS: Okay, okay. Any other questions before we go to voting? Okay. Larry and Bob.

MEMBER MILLER: I'm sorry. Just a quick technical question. Would this come back in three years like normal things or just when would this come back in terms of arguing this again?

DR. BURSTIN: It has to come back within three years. It doesn't have to wait three years. As soon as they're done with their testing and their evaluation, we hopefully get it back much sooner than then. We don't want to wait that long if we don't have to.

CO-CHAIR PINCUS: Okay, Bob.

MEMBER ATKINS: The unpackability question. If somebody screens positive on two or three of the substance classes, will the counseling be one counseling for the complete set that they score a positive on? Or do they need three separate counseling sessions?

MR. WALLER: They would all need to be addressed, but it could be in one session.

MEMBER ATKINS: Okay. I don't know what that has to do -- it seems to me that might be complex, in terms of the unpacking. But you guys deal with that.

CO-CHAIR PINCUS: Yes, okay. So why don't we move ahead? So okay. So we're going to vote now on this issue about the evidence for the importance of the measure of concept, and just to clarify, that a 1, 2 or 5 would allow the measure to go forward and for us to review the other criteria. Is that correct, Helen? Okay.

MS. DORIAN: Caroline, are you on the phone?

(No response.)

MS. DORIAN: All right. Does everybody have their clicker and ready to vote? Okay.

MS. BAL: Voting is now open for evidence. The options are 1 high, 2 moderate, 3 low, 4 insufficient evidence, 5 insufficient evidence with exception, and as we said earlier, 1, 2 and 5 combined together would allow the measure to pass. One more.

CO-CHAIR PINCUS: Les isn't here.

MS. BAL: Is anyone else out of the room? We're missing one person on that side. Let's move forward with 22 then.

Okay. So the results for 2597 evidence is high 3, moderate 5, low 0, insufficient 1, insufficient evidence with exception 13 and we'll move forward to gap.

CO-CHAIR PINCUS: Okay. So let's move forward, and Peter.

CO-CHAIR BRISS: Essentially everybody agreed that there was a large performance gap.

CO-CHAIR PINCUS: Okay.

MS. BAL: Okay. Voting for gap is now open.

[VOTING.]

MS. BAL: Okay. The results for gap for 2597 is high 17, moderate 5, low 0, insufficient 0, and we'll move forward to high priority. Would you like to discuss or just vote?

CO-CHAIR BRISS: No issues on priority.

MS. BAL: Okay. Voting is now open.

[VOTING.]

MS. BAL: Okay. The vote for 2597, high priority, is high 20, moderate 2, low 0, insufficient 0, and we can also vote for the composite importance as a whole. Would we like to do any discussion or just

move forward with it?

CO-CHAIR BRISS: So it was articulated, and I think we've discussed the issues with the composite already.

MS. BAL: Okay. Voting is now open.

[VOTING.]

CO-CHAIR PINCUS: By the way Helen, you may want to modify this criterion for the purposes of e-measure testing, okay.

(Pause.)

MS. BAL: We're just missing one vote. Okay, we got it. Thank you. Okay. So for the importance composite, for 2597, we have high 8, moderate 9, low 3, insufficient 2, and we can move forward to the trial use specifications.

CO-CHAIR BRISS: So essentially the work group thought that it was cleanly specified.

CO-CHAIR PINCUS: Are there other comments with regard to the measure

specification?

(No response.)

CO-CHAIR PINCUS: I guess, just -- I think there was some concern that Raquel had brought up about the specification of the counseling component.

CO-CHAIR BRISS: But as with many of the measures yesterday, you know, we talked about it. Sometimes we've essentially accepted a developer's argument that they're trying to be flexible, and I think that we've heard that argument again today.

CO-CHAIR PINCUS: Tami.

MEMBER MARK: I thought I heard that the specification is going to allow for easy billing. I just want to confirm that, because it seems important to, you know, to encourage that.

CO-CHAIR PINCUS: Could you speak a little closer to the microphone?

MEMBER MARK: I thought I heard that the specifications that allow for easy

billing, and I just want to confirm that, because it seems like adoption of the screening is going to be more encouraged if the specification is linked easily to billing.

MS. HANLEY: As long as the counseling services provided meet the criteria to bill for that, it can absolutely be used for the measure. But we're not measuring whether or not what's right to bill or not. So I'm clear by what you mean by easy billing.

MEMBER MARK: Well, I mean maybe you could just articulate a little bit how like checking the boxes would be translated in a provider organization into something that would be billable.

MR. WALLER: Yeah. So basically you have three levels of brief intervention billing, and those are connected to the specification when it's started. So I mean you go through the screening if it's positive. Then at the end, on your billing screen on the EMR, you can bill either -- you can put non-

billable brief intervention which is the really short version, like you chatted with it but it's not something that needs billing, or you can actually code it out as the specific E&M code for each of those two.

And it's literally the way that it's built into the EMR, check which one applies. Did you do extensive, you know, and it's based on time. So it's based on 15 minutes is billable. If you do less than 15 minutes, it's not billable.

Now that 15 minutes can be combined. So if the physician does a portion and another portion is done by their social worker or nurse, that can be combined billing, and that meets specs for reimbursement. So it can be billed out, even if the physician does half, and if the other half is done as team, then that can still be billed out so --

MEMBER MARK: So is the screening part billable or only if the screening leads to brief intervention is it billable?

MR. WALLER: Screening is also a billable element. Whether or not it's paid is, you know, built on different payors. There is a code that can be levied for the screen.

(Off mic comment.)

CO-CHAIR PINCUS: Okay. So are we ready to vote on specifications?

MS. BAL: Okay. The voting is now open.

[VOTING.]

MS. BAL: Just waiting for two more. If everyone could please vote.

(Pause.)

MS. BAL: So the trial measure specification for 2597 is high 4, moderate 16, low 0, insufficient 2, and we can move forward to feasibility.

CO-CHAIR BRISS: Okay. So there are feasibility questions about this measure, which I think we've already discussed.

MS. HANLEY: Just so everybody is

aware, we did do feasibility testing and two sites, a federally qualified health care network of clinics, and a solo practitioner family practice site, and that was included in your submission materials with the feasibility assessment.

CO-CHAIR BRISS: And it does raise some questions though, because the solo practitioner scored high, and he or she was presumably motivated. The community health center didn't implement it very well, as I recall.

CO-CHAIR PINCUS: Okay. Any other comments?

(No response.)

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

MS. BAL: Okay. Voting is now open for feasibility.

[VOTING.]

MS. BAL: Okay. The results for 2597 feasibility is high 4, moderate 14, low

4, insufficient 0, and we'll move forward to usability and use.

CO-CHAIR BRISS: I don't think there's anything to talk about here that we haven't already talked about.

CO-CHAIR PINCUS: Okay. Any other comments?

(No response.)

CO-CHAIR PINCUS: Okay. Let's move to voting.

MS. BAL: Okay, voting is now open.

[VOTING.]

MS. BAL: Okay. The results for 2597 usability and use, high 10, moderate 9, low 1, insufficient 2, and we'll move forward to the overall vote, unless you would like further discussion.

(Off mic comments.)

MS. BAL: Do you want more discussion or no?

CO-CHAIR PINCUS: Any further

discussion?

(No response.)

CO-CHAIR PINCUS: Okay.

MS. BAL: Voting is now open.

[VOTING.]

MS. BAL: Okay. So for overall suitability to be a trial measure for 2597, the results are yes 20, no 2. So this measure will be a trial measure -- trial use measure.

MR. WALLER: I'd like to thank the Committee for their time in looking at this. I know it was confusing, so thank you.

CO-CHAIR PINCUS: Okay. So thank you very much. I think Helen, we were glad to be a guinea pig for NQF on this process. I think we learned a lot. I think it's interesting. I think there's a -- you know, I think there's a reality that, you know, almost by definition, you have to apply different standards, because you're trying to get information.

So it really is not looking out

there that, you know, that whether or not this should be used, but really will this generate information. I guess a couple of just recommendations to the developer, just a couple of things, just comments.

I received an email from Rich Sates, sort of raising a series of issues, raising a series of issues, and he had asked me to bring it up. I didn't do it, because it's clear that he, not being part of this interaction, wasn't completely informed about all the issues. I suggest you might want to contact him to --

(Off mic comment.)

CO-CHAIR PINCUS: And the second thing is -- the second thing is NIDA had a recent meeting that Connie and I were both at, in which they are sort of actually developing these common data elements as part of their clinical trials network, and actually looking at sort of specifying measures, and it would be good to coordinate with them.

Okay, good. So now we're going to move to Measure 2599, and Peter's going to chair and Mady's going to be --

CO-CHAIR BRISS: So this is alcohol screening and follow-up. It's another, and would the developer like to pick this up?

MS. DORIAN: So just to note, we're not sticking with the original agenda. 2605 was up next, but we're going to do 2599 and 2600, and then 2605.

MS. HUDSON SCHOLLE: Good morning.

So we're continuing on our journey to look at care for people with serious mental illness. So this morning, there are two more measures that we're going to review, that address alcohol and tobacco use, and so these are similar to the BMI measure that we discussed yesterday.

So again, these measures looked at existing measures from the existing NQF-endorsed measures. All three of the measures,

BMI screening, alcohol screening and tobacco screening, all three of those measures are set up as screening and follow-up.

So this is not about full treatment. This is about getting people connected and doing the first step in follow-up of a positive screen. The alcohol measure focuses on people with serious mental illness.

So the two changes that we made were we took that existing provider level measure, that looked for screening and immediate follow-up at the time of the screening, and we adjusted it for reporting by health plans for the SMI population.

So the denominator is now people with serious mental illness, and the numerator is looking for two events of counseling. Now there's a question about what should count as meeting this numerator, and so we allow two events of counseling.

There was a question, I think, whether medication use would be consistent for

alcohol. We did not -- I'm sorry, and but that was not included in the original measure, which focused on counseling, and so it's not included here.

The counseling can occur at the visit by the provider who conducted the screening, or it can be other kinds of events. We increased the expectations to two events, given that this is a high risk population and this single event was -- our panel thought that a single event was not sufficient.

Okay. I'm sorry. I'm speaking right into the mic. Okay, okay, exactly. So the numerator requires two events of counseling. It can be conducted at the visit at the time of the screening, or it can be -- and those two events have to occur within three months.

They can be done by any kind of provider, including a health plan care manager, if the health plan can document that follow-up. I think that's where I'll stop.

I mean we based this on evidence about the prevalence of alcohol use, and this was tested in a way that was similar, in the same sets of health plans.

I will note that we were -- that people with an existing alcohol use disorder are excluded from this denominator because they have already been identified. This is about identifying new cases. So a number of people in this population are already identified, so that decreases the denominator.

It also -- what we found in our testing was a very low rate of positive results, which our stakeholder group said doesn't make any sense at all. So we were -- that's what we found.

Now we largely had access only to medical records. So it's likely that the alcohol use might have been identified and addressed in behavioral health care. But as we talked about yesterday, even though we were in places where we -- theoretically where the

plan had access and their nurses were able -- the plan requested behavioral health records, they did not always get those in time for this review.

MS. LIU: I'd just add that these measures are to address the comorbid condition. So you see that for the SMI population, we try to address their alcohol use, screening and follow-up, and also tobacco use for both SMI and the AOD population.

We also reviewed these measures with the original measure developers and stewards, and they felt our adaptation of these measures made sense, given that these are applying to another vulnerable population. The results, as Sarah mentioned, demonstrated that there was a gap in care and there was disparity.

CO-CHAIR BRISS: So let's let Mady key this one up for the Committee, and then we'll have a discussion.

MEMBER CHALK: Clearly, there's a

performance gap that was identified by the measure developers. However, the data are very limited, because the alcohol screening and brief counseling measure has only been recently endorsed by NQF. So it hasn't been terribly well used out there.

No issue with regard to reliability. The issue of validity remains, as we talked about some yesterday. But in this one, the measure -- I still have issues with the fact that the measure was not tested in commercial health plans, but was tested only in a variety of Medicaid and Medicare plans, and that the data was shockingly bad.

I mean it's much worse than the data that we saw about the measures yesterday. I mean it's almost -- there was one -- one of the plans only screened one person. So there's, you know, I don't -- it is of big concern to me that we push, we endorse a measure like this with those kinds of issues and lack of testing.

No issues with feasibility, because alcohol claims are not difficult to identify and reference. Usability, as the first reviewer, this is coming just from me.

I have, as I said yesterday, some major concerns about pushing forward quality measures such as this, that come with no action plan to help them be implemented appropriately and useful and move, related to improving quality.

I don't have any expectations personally that this measure will do anything more than has happened with the screening measures for alcohol and -- well, for alcohol, that we have -- that NCQA has currently, which haven't moved at all. So you can take that for what it's worth.

There are significant disparities in the care that was provided, but NCQA talked about that a minute ago. Only a third of the people, about a third of the people who were screened have follow-up in the three Medicaid

plans, despite the fact that we know that alcohol is heavily implicated in major chronic illnesses for people with serious mental illness.

There was good face validity. However, there is one issue that was of deep concern to me, and that has to do with the fact that the two events that NCQA just talked about of counseling, the specifications were amended to allow self-help services to be documented in the clinical record, which means AA, and to this reviewer, that's unacceptable.

MEMBER MAZON JEFFERS: So I'm just trying to, in my head, reconcile three things: The intended use of the measure at the health plan level, the care setting that people with SMI are normally treated in, and the evidence of effectiveness for the SBI in that SMI population.

So I think I understand that the intended use is at the health plan level, and that people with SMI would be screened in a

primary care setting for the alcohol use.

My understanding of a normal clinical pathway for someone with SMI is that I don't know how much contact they have with their primary care physician, and whether the care setting, screening them for their alcohol use makes sense to happen in the behavioral health setting, or in the -- in the primary care setting.

I guess you're going to tell me it could be used in either, but it's a little confusing, and then the third thing is that as was discussed before, screening and brief intervention is more effective for alcohol use, in terms of changing outcomes, in a population that has mild to moderate substance use.

The SMI population really might not fall into that category. It might not be the most effective population to engage using SBI.

CO-CHAIR BRISS: I'd like to focus

us on what -- try to focus us on one issue at a time. So if we -- if for now we can talk about the evidence things, we'll come back to the rest of the criteria.

MEMBER SHEA: This is just a brief question. Just looking at the specifications, one of the screeners was the CAGE, but I know we had a large discussion about how the CAGE really wasn't a good screener. It's more for someone who has dependence, and so I was just wondering about that.

MEMBER TRANGLE: This is more of a technical question. I mean we've been doing this, and my own perspective is as an integrated system, we're trying to do it at the plan level telephonically. We're trying to do it in primary care and behavioral health.

I heard you say a plan phone call well documented counts. Is that also true if it's a nurse or care manager in primary care/behavioral health? It's documented but

they're not a billable provider? Is it coming from the EMR or is it claims-based I guess is what I'm asking? Could it be both?

MS. HUDSON SCHOLLE: Okay. So it is plan level. So that's right. It counts. It doesn't have to be a claim. It could be either a claim or documentation in the medical record, or documentation, wherever the health plan might document that case management service. So I think what you suggested about the nurse, that would count.

MEMBER TRANGLE: Okay. As long as it's documented in the EMR, it can be abstracted somehow.

MS. HUDSON SCHOLLE: Right.

MEMBER TRANGLE: Whether there's a billable provider or not.

MS. HUDSON SCHOLLE: Right, right. So we're looking for the documentation or the screening and the service, okay. So there's not a requirement about who does it. The second thing is it is intended to be wherever

it happens, right?

It could be in primary care or it could be in the behavioral health setting. It's just to make sure that it actually is documented and addressed. The third question is about which, the CAGE or the audit or what actually counts.

So in the original measure that we're building on, anything counts essentially. The CAGE or the auditor both listed, as well as documenting frequency and amount of alcohol use. So it's basically documentation of any kind of alcohol use.

In terms of is this the right intervention for this population, I think this is something that our Committee struggled with, and again, this measure is mostly about trying to make sure people get on the pathway to treatment, okay.

So counseling and it's based on the SBIRT logic. But the panel felt very strongly that we needed to capture efforts to

connect people to peer support, peer-lead interventions, as well as other interventions. That's why this has been a constant criticism of our existing measures that look at initiation, engagement and treatment, and concern about that.

Because this is about connecting people to service rather than evaluating the completeness of their treatment, then that's why the panel felt like this measure of screening and making an active step towards follow-up, that that was a reasonable thing.

The issue is to have it documented in a record where it could be identified, and that was a strong recommendation from the consumer members of our panels.

MEMBER ZUN: So as I read the numerator, it says receive two events of counseling to identify alcohol abuse; is that correct? And so the problem I have with that is we know that many problem drinkers are recalcitrant to follow-up and treatment. But

then you're holding the plans accountable for ensuring that the patient actually gets that.

So you know -- so from my perspective, you know, you can lead them to the paper but you can't make them drink. So if we hold them accountable for leading to the water, and not necessarily making them drink. So we can't control them actually going to those appointments or going to AA, or getting any intervention.

The responsibility of the plans are to make the referral, to connect them. But whether they go or not, you know, how do we enforce that with anyone?

MEMBER MILLER: I think this is an incredibly important issue, the co-occurring disorders. We were talking about it last night while we were drinking wine, the work I've done to really pull this together in Arkansas, and we've done that to some extent.

My concern is, following up on Raquel's question about the -- and at the CMHC

level, for instance, Medicaid in our state does not pay for alcohol treatment. So even if there is the screening, there's no incentive to do it. There's no incentive to do it because no one pays for it, and I'm sure there are other plans like that as well.

Now in the future we will be, hopefully starting in January, if the legislature doesn't push that back. But I'm concerned about the operationalizing of this and the evidence that this really going to happen, and is there incentive to do it because there's no process for paying for it? I think that's an incredibly important consideration that we found already.

MEMBER ROBINSON: Last question I can ask, I guess.

CO-CHAIR BRISS: That's right. This is your last chance. Make it a good one.

MEMBER ROBINSON: I'm off the island, okay. Okay. So obviously there's a huge problem if the testing only revealed a

very small amount of screening that's going on, and there were substantial research done in the past that shows 60 to 70 percent of those individuals with SPMI that were hospitalized had some type of alcohol and/or drug use.

So we're definitely missing the boat with this. I'm just not clear, number one, several things. With this being a health plan measure, I agree with the gentleman down the table. Sorry, I can't see your name tag from here.

Okay, Larry, that this could be a very difficult measure for a health plan and would be more appropriate for a health system. Whether it's a facility inpatient, an ER or a primary care practice or outpatient behavioral health practice, primarily from the standpoint of you that -- and I say that from the standpoint of you, the direct motivation, if the way of being able to apply this is clear.

I agree with the gentleman down

the table, that there's so much variation amongst the states as to how substance use disorders are paid for, that it really becomes real complicated to even implement, because of the barriers just from a benefit perspective.

So I guess those are my concerns about it. My question is is the counseling that can be counted, are there specific codes for this that would help differentiate the counseling from any type of psychotherapy that this population would naturally receive, and would this also require in the specifications the use of a diagnosis if there is a positive screen, because that's another issue that is problematic in this regard.

CO-CHAIR BRISS: So let's take a second and let you make some responses.

MS. HUDSON SCHOLLE: Okay. So the last question, does it require a diagnosis? No, okay. There's nothing where we say that. It's just a positive screen. So we're not looking at any diagnostic information. We're

really looking at the services codes and all of that could be billed in the medical --

I mean it could be described in the record or in the care management record, without actually providing a diagnosis of alcohol use disorder, right, because this is for unhealthy alcohol use, not -- maybe not reaching the disorder criterion.

The second question or previous question was about who's accountable? Where's the right level of accountability, and so, you know, in our experience, primary care providers hate measures like this, that require them to do things that -- where they are concerned about whether they have either the expertise on staff or a connection to the specialty services that their patients might need.

And in particular for this population of people with serious mental illness, what we're looking at is something that actually should happen in behavioral

health. If it doesn't happen in behavioral health, you want it to happen in primary care. You want people to be thinking about the comorbid condition.

So that's why, in our view, having the health plan responsible says health plan, make it happen and make it happen wherever it makes sense within your network of providers. So and I -- and we are cognizant of the challenges of the fragmentation of services for mental health, substance abuse and general medical care, and who pays for it where.

I can't -- we started with 50 plans that wanted to test these measures. We found three that were still able to test the measures, because we had -- we said you had to be responsible, and we went through a lot of states where the states carved things out, and we had plans willing to test it, particularly plans that were serving dual eligibles, where Medicare expects some things and then the Medicaid program might carve things out.

So if you think about this whole suite of measures, they're really assuming or arguing for coordination and integration of care, and so in an -- so they assume, they require the sharing of information across all those diverse settings, and they put the accountability at the health plan, because the health plan may be paying for, you know, multiple pieces of that.

It could be that the state is really, and the state and Medicare and commercial payors as well, that create this kind of fragmentation, these measures, if they want these measures, they're going to have to think about well how do we make this happen?

So you know, just saying that those systems don't talk and we're not accountable, yes and people with SMI really should be screened and get some follow-up if they have an alcohol problem.

CO-CHAIR BRISS: So if possible, I'd like to quickly come back up this way, and

then vote on the evidence, please. Okay, Jeff.

MEMBER SUSMAN: So just a very brief comment, and mostly in reaction to Les. I think we've gone beyond being able to say oh, it's our job to lead the horse to water, but not to make them drink.

I think the developer captured that in her last remarks around, you know, at a health plan or at system level, I think we need to figure out ways to get the horses to drink.

It is our responsibility, and we used to say that about a lot of things that we've been able to make substantial improvement, progress on. So while we may do a really awful job now and have a single person, I think that is all the more reason to go down this path and support development in this arena, to provide the data and make it compelling, and to not only lead horses to water, but find ways to creatively get them to

drink.

CO-CHAIR BRISS: Michael. Stop drinking, yeah. Or substitute water for whatever else they want to drink perhaps.

MEMBER TRANGLE: So unlike Jesus, it's not going from water to wine. We're going to the other way. But I want to piggyback on that, in that as a system that's been doing this and trying to do this better, we do SBIRT in the hospital.

We do it at different places. You know, not perfectly obviously, but a lot of people, you know, it's not -- we're in a world where it's not a total closed system. We're not a Kaiser.

So especially in a CD, people go all over the places for where they're going to get their treatment or not, and to have a measure that ideally -- it would be ideal for this measure to span health plans and the joint commission, you know, so that in some sense, both could be requiring it.

But it's a way of compiling data, so that everybody has to share their data and kind of work together, to find out where they're going and what's happening. If you leave it splintered, you know, with the hospital world not thinking about it, the plan world thinking about it and whatever's going on in primary care and behavioral health, segmented off.

The more we could have these measures cross all systems, the more we're going to learn, the more we'll be forced to sort of improve what's happening for patients in a more coordinated way.

CO-CHAIR BRISS: So I hope we can go through Bob and Mady relatively quickly, and then try to get to a vote, please.

MEMBER ATKINS: So every one of our health plans, our business rules, if someone's positive for mental illness, we screen them for substance use. Because of all the things you talked about, clinically,

passionately I agree with you.

I think you have two measures here, not one, and they have different kinds of issues involved. I think there's a to what extent do people get screened, which is a really important question, and then do you do something about it, and the something --

Because we have plan states that don't cover it, because it's program funded through the Department of Addiction Services. But then our care managers are supposed to coordinate outside of Medicaid to make it happen.

So I think I haven't read in this that you can capture effectively the work that's done for the second metric, which is have you done something to hook them up with the service, and if I have it in a progress note in my computer system, I'm not sure that the metric will very effectively pick that up. So I think it's two metrics.

CO-CHAIR BRISS: So that issue

might be -- that issue might be more about reliability and validity than importance to measure. So Mady, will you finish us up please, and then let's go to a vote?

MEMBER CHALK: Okay. So while I agree with Michael about the importance, I still have to ask the same question I asked before, which is does follow up for people on Body Mass Index and screening, is pure support good enough to count as counseling?

CO-CHAIR BRISS: So is the question you're asking, Mady --

MEMBER CHALK: Is it the same standard that you're using for the alcohol and SMI measure or not?

CO-CHAIR BRISS: So is the issue that there's really not an obesity-related AA?

MEMBER CHALK: No, it doesn't have -- well, there is as it turns out. There is as it turns out. Overeaters Anonymous, and it's across the United States, that group.

I want to know if that's adequate

care for somebody with serious mental illness, and are you saying that for alcohol it is adequate care, but it's not adequate for people who are obese?

CO-CHAIR BRISS: So do you want to make any other comments or extensions? Fine.

(Off mic comments.)

MS. HUDSON SCHOLLE: Our panel recommended that we include it in this step, because they felt like if four, that this would -- that that is a reasonable follow-up step for people with healthy alcohol use. Not substance use disorder. We have a separate measure that looks at the substance use disorder and engagement and treatment.

MEMBER CHALK: Body mass index is not a disorder. It doesn't show up.

CO-CHAIR BRISS: So let's -- this has been a spirited discussion. Let's try moving through the voting on importance to measure and report please.

MS. BAL: Voting for evidence is

now open.

[VOTING.]

MS. BAL: We're only at 20 and we need three more votes. If everyone could just vote please?

Never mind then. All right. So for evidence for 2599, we have high 7, moderate 11, low 1, insufficient evidence 1 and sufficient evidence with exception 1. So we'll go forward to gap, and do we want to discuss or just start voting for all the importance ones?

CO-CHAIR BRISS: Does any -- I think we've likely talked about this already. Does anybody feel an urgent need to say anything else that hasn't been said?

(No response.)

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

MS. BAL: Okay. Gap is now open.

[VOTING.]

MS. BAL: Okay. So for gap for

2599, we have high 17, moderate 3, low 1, insufficient 0, and we'll move forward to high priority. Voting is now open.

[VOTING.]

MS. BAL: Okay. The results for high priority for 2599 is high 18, moderate 3, low 0, insufficient 0. Now we can move forward to scientific acceptability.

CO-CHAIR BRISS: So it seems to me that we've opened many of these issues already. Are there other comments on reliability or validity that haven't already been made that folks would like to address?

MS. HUDSON SCHOLLE: I just want to clarify that this is set up as a what we call hybrid measure, right, where we identify the denominator from claims data, and then the chart, the numerator can be identified either from claims data or from medical records review or other supplemental data that the health plan has.

So the progress note discussion

documentation, right, would count or a nurse in the clinic. But it has to happen twice. So you can't just say we referred them to the other clinic. It has to say "referred," and then there's a visit to the clinic.

(Off mic comment.)

MS. HUDSON SCHOLLE: Would allow it to -- the hybrid methodology would allow the plan to review medical records.

(Off mic comments.)

MEMBER ATKINS: I'm not the expert at hybrid methodology, but I've always understood it means we send people out to providers, and look at their charts. I didn't know that we could do an internal review of our internal plan records as part of hybrid methodology. So maybe I just learned something. I just didn't know that.

MS. HUDSON SCHOLLE: Trust me, your plans know how to do this.

CO-CHAIR BRISS: So Mike.

MEMBER LARDIERI: I'm just

confused. So how do you document the peer, because peers don't have a record?

MS. HUDSON SCHOLLE: It would have to be documented in the record, saying not that the person was referred to AA, but "person reports that they are attending AA."

MEMBER LARDIERI: Oh, okay.

CO-CHAIR BRISS: So other comments on reliability and validity?

(No response.)

CO-CHAIR BRISS: Let's move to voting, please.

MS. BAL: Okay. Voting for reliability is now open.

[VOTING.]

MS. BAL: So the final result is high 0, moderate 12, low 8, insufficient 1 for 2599, and that will actually put us in the gray zone, but we'll move forward, and voting -- unless you want to speak further.

CO-CHAIR BRISS: And any final arguments on validity before we vote?

(No response.)

CO-CHAIR BRISS: Hearing none.

MS. BAL: Okay. Voting for validity is now open.

[VOTING.]

MS. BAL: So the result for validity is 2599 is high 2, moderate 10, low 6, insufficient 3. So this will also be in the gray zone. We can move forward to discussion of feasibility.

CO-CHAIR BRISS: Any new issues on feasibility that haven't already been raised?

(No response.)

CO-CHAIR BRISS: Hearing none, let's move to voting please.

MS. BAL: Okay. So voting is now open for feasibility for 2599.

[VOTING.]

MS. BAL: Okay. So the results for feasibility for 2599 is high 1, moderate 11, low 8, insufficient 1, and it's also in the gray zone, and we'll move forward to

usability and use.

CO-CHAIR BRISS: Any final arguments on usability and use before we vote?

(No response.)

CO-CHAIR BRISS: Hearing none, I think we can move to vote.

MS. BAL: The voting is now open.

[VOTING.]

MS. BAL: Okay. The results for usability and use for 2599 is high 2, moderate 12, low 5, insufficient 2, and we'll actually pass that criteria, and then we can do overall, unless you would like further discussion.

CO-CHAIR BRISS: And any closing arguments before we talk about -- before we vote on overall suitability?

(No response.)

CO-CHAIR BRISS: You do vote if it's in the gray zone. You only completely reject a measure if it's under 40, not if it's gray.

MEMBER PATING: Just a question about the process. So what does that mean though? Did staff work on it or is there improvements in reliability and validity or how do you --

CO-CHAIR BRISS: I think the quick answer, the quick answer is in the spirit of transparency for the stuff going forward, we highlight more than usual this agreement. But it otherwise gets treated in much the same way as if we had fully approved it. Is that fair?

MEMBER PATING: Because yesterday we had two that were actually -- I thought we voted up on, but then we're reporting out that they were gray zoned. There was one?

CO-CHAIR BRISS: There was one.

DR. BURSTIN: One from yesterday. Yes. So essentially it's just again, as Peter said, it's about transparency. These measures will go out with a clear indication of which criteria had gray zone or overall. So it just invites comments specifically on the issues

you've already raised.

MEMBER PATING: If we vote up, that means that we're letting it go through then? There's no next process.

DR. BURSTIN: It will still be in the report with gray zone for the criteria you reported it as for gray zone, so people will know you had specific issues in these criteria.

CO-CHAIR BRISS: And recall that there are other steps after ours. So this flags areas where the Committee had more than usual disagreements. So that people like the CSAC and the member councils and the other people that look at these measures and consider them can know what our sticking points were about the measure.

(Off mic comment.)

CO-CHAIR BRISS: So this will be labeled as something like 50 shades of gray, right, and so overall suitability. I think we're probably ready to vote.

MS. BAL: Okay. Voting is now open.

[VOTING.]

MS. BAL: There's one vote out there somewhere. Oh, there we go. Okay. So we have yes 13, no 8 for overall suitability for 2599, and that is just passing for endorsement.

CO-CHAIR BRISS: So thank you. I think what we'll do next is we've got two more measures to do before the lunch break. I'd like to take a ten minute break. Please be back in your chairs by right on 11:00 a.m. We'll get restarted and we'll have two more measures to do by noon. So we'll have to be efficient after the break.

(Whereupon, the above-entitled matter went off the record at 10:46 a.m. and resumed at 10:57 a.m.)

CO-CHAIR BRISS: So the next measure is the tobacco screening and follow-up measure, and Susan, will you please kick us

off.

MS. HUDSON SCHOLLE: Sure. Okay. So this measure, like the alcohol and the BMI measure, this is looking at tobacco use screening and follow-up. Again, it's adapted from the provider level measure.

This one does -- follow-up for tobacco use includes either medication or counseling or both -- and two events. The two events can be any of those.

I did want to emphasize, you know, this measure, we do have good inter-rater reliability and trying to get this information from the records, and we see A, a gap in care and disparities, and certainly what we heard from our panel is that they -- in the stakeholder focus groups, people thought this was a truly important topic that ---- we asked them to prioritize things and they put this one very high.

CO-CHAIR BRISS: So the lead discussant for this one is Constance. No?

FEMALE PARTICIPANT: No, it's Mady. No, I'm sorry ---- Kraig Knudsen.

MEMBER KNUDSEN: I think it's me. There's a conflict with the other individual. So it's all about me. So in terms of evidence, obviously there is plenty of evidence on the harmful effects of tobacco usage. Individuals with mental illness are more likely than the general population to smoke, 31 percent versus 21 percent.

Obviously that puts them at risk of other conditions. Study assessing tobacco use among individuals with substance abuse treatment settings showed a prevalence of smoking, around 77 and 94 percent. Obviously tobacco use is related to poor health outcomes.

In terms of work group comments on evidence, the Committee agreed that it examines a critical issue and is important. However, it did stress the fact that evidence indicates pharmacotherapy for smoking is most

effective when it is including counseling. So that's important.

Additionally they had concerns, specifically with the SMI population, that adding another medication may not be the best approach, and also that the definition of follow-up care for this measure did not require both pharmacotherapy and cessation counseling. So that's that with the evidence.

CO-CHAIR BRISS: So this is sort of --- this is very -- obviously, this is very conceptually related to the last measure, and so some of the issues will have been -- will likely have been raised. It's always hard to argue that tobacco screening and treatment isn't an important thing to do, right?

So are there -- I suspect we might be able to keep this discussion fairly brief. Are there additional comments that folks would like to make about the importance to measure and report on this one? Raquel.

MEMBER MAZON JEFFERS: Could the developer just address the issue of why you made the measure pharmacotherapy or cessation counseling and not both?

MEMBER SHEA: So actually we had -- the specific issue is that in thinking about the SMI population. So this measure is for both the SMI population and AOD population, and people were concerned about requiring that people with SMI take another medication.

And so they felt like allowing medication or counseling to meet the measure numerator was important to allow the patients to have the choice of which approach.

CO-CHAIR BRISS: Any other -- any other questions or comments on importance to measure and report?

Hearing none, why don't we try to move to voting, please.

MS. BAL: Okay. Voting is now open for evidence for 2600.

Okay. So for evidence for 2600, is high, 18. Moderate, two. Low, one. Insufficient, zero. Insufficient with exception, zero, and we can move forward to gap, unless there's further discussion.

CO-CHAIR BRISS: So anybody want to make further comments about a performance gap?

MEMBER KNUDSEN: In terms of the workgroup comments and also what was presented, they did a study that showed 35.8 percent of people with severe mental illness had tobacco use screening and appropriate follow-up, which I think is pretty alarming, and for alcohol and other drug dependence population, the field tests showed that 22 percent had tobacco screening and appropriate follow-up, which I also think is absolutely alarming.

There's limited data on disparities, and they were not able to assess differences by race and ethnicity or language.

So that's it.

CO-CHAIR BRISS: So we have a vote for absolutely alarming? Anybody like to -- anybody like to -- it's a call and raise over absolutely alarming. Hearing none, why don't we try to move to vote.

MS. BAL: Okay. Voting is now open for gap of 2600.

CO-CHAIR BRISS: And the truth is, for the NQF staff, I think there needs to be an actual category for absolutely alarming. Could we suggest that as a methodologic improvement?

MS. BAL: Okay. So the results for gap for 2600 is high, 18. Moderate, one. Low, zero. Insufficient, two, and we'll move forward to high priority.

MEMBER KNUDSEN: Everybody thought it was a high priority.

CO-CHAIR BRISS: Any further comments? All right. Then we'll vote please.

MS. BAL: Okay. Voting is open

for high priority.

We're just waiting for one more person. If you'll retry it please. So the result for 2600 high priority is high, 16. Moderate, four. Low, one. Insufficient, zero, and we can move forward to the discussion on scientific acceptability.

CO-CHAIR BRISS: So would you like to tee it up for us?

MEMBER KNUDSEN: All right. On this one, they did an inter-rater reliability test, and the final sample, it included -- I believe, let's see here, a few health plans. The final sample is 756 patients with SMI, and 306 patients with AOD. High inter-rater reliability, and then in terms of validity, they did their face validity with an expert panel of 16 folks, group of 29 and public comment of 20, and it was found to be valid.

So the kappa was .75, and the overall .57. So there were some other comments on this. High rates of missing

records present, presented a challenge for the generalizability of the population. Only about a third of the patients had behavioral health records available.

A Committee member expressed that the pediatric population should be included, rather than limiting the measure to those over 18, and the Committee challenged the limitation of the measure to outpatient settings, noting that much care is now delivered in acute settings. So that's what they talked about.

CO-CHAIR BRISS: An issue -- I think this Committee's talked about this before. There's a general issue with tobacco measures in adults and kids. This kind of measure, where part of the specification is drug treatment, because there's sort of insufficient evidence to conclude on whether drug treatment for tobacco actually works in kids.

So any other comments on

reliability before we move to a vote on this one?

MS. BAL: Okay. Voting is now open for reliability.

Okay. So we have ---- for reliability of 2600, we have high, six. Moderate, 14. Low, one. Insufficient, zero, and we can move forward to validity, unless you guys would like to discuss more.

MEMBER KNUDSEN: I already discussed what the -- any other comments?

CO-CHAIR BRISS: Any other comments? No.

MS. BAL: Okay. Validity is now open.

Okay. So the results for validity for 2600 is high, five. Moderate, 14. Low, three. Insufficient, zero, and we can move forward to feasibility.

MEMBER KNUDSEN: Everybody thought it was feasible.

CO-CHAIR BRISS: That was

admirable efficiency. Any other comments? All right. Voting please.

MS. BAL: Okay. Voting is open for feasibility.

Okay. The result for feasibility is high, seven. Moderate, 12. Low, two, for 2600, and we can move forward to usability and use.

MEMBER KNUDSEN: There were no comments on usability.

CO-CHAIR BRISS: Okay. Would anyone like to add a comment on usability? Seeing no cards, let's vote.

MS. BAL: Okay. Voting for usability and use is now open.

So the results for usability and use for 2600 is high, six. Moderate, 14. Low, one. Insufficient, zero, and we can move forward to overall vote, unless there's further discussion.

CO-CHAIR BRISS: Any closing arguments? No. Let's vote, please.

MS. BAL: Okay. Voting is now open.

CO-CHAIR BRISS: Oh, I'm sorry. All right, I see your card. Dodi.

MS. BAL: Not open.

MEMBER KELLEHER: I just wanted before we ended --

CO-CHAIR BRISS: Please use your mic.

MEMBER KELLEHER: Oh, microphone, sorry. I've got it. I just wanted to -- because this has come up with other phases as well.

With an SMI population like this, you really ought to consider, to increase your validity, sampling from other than just outpatient, because so many of that, especially the Medicaid population, do get screened and treated perhaps more reliably in acute and subacute levels of care, such as residential or, you know, subacute locked, which are quite common.

So I think you need to go back and look at that the next time around. That's all. It's not really an argument for yes or no.

MS. HUDSON SCHOLLE: Okay. So inpatient would not count, but I believe that intensive outpatient would count like at the intensive outpatient settings. No?

MEMBER KELLEHER: Okay, go for it.

MEMBER TRANGLE: I mean health plans pay for inpatient, straight outpatient or IOP services. But a lot of these patients go to group homes. Some Borden Cares, ACT teams, you know, a lot of their treatment, including where they reside, is in a psychosocial sphere versus the medical sphere.

That's where they're living, and you can do a lot of really important key care there that we've never thought of as a gap in our conceptual continuum of care -- of who owns it, you know.

MS. HUDSON SCHOLLE: So if the health plan could track that that service --

health plans typically do not look --

MEMBER KELLEHER: No.

MS. HUDSON SCHOLLE: They don't -- they typically don't look at care that's provided inpatient, but those services that if they're provided -- if the health plan can produce the record that demonstrates that it's done, this is a hybrid measure. If they could find it, that it's happening there, it would count.

MEMBER KELLEHER: You know, if you're looking at -- if you're looking at quote-unquote medical record or notes, for say someone might be -- either for SMI or for substance abuse, might be in a 30 day treatment residential that's paid for by the health plan, and they're much more likely to be assessed and given interventions for things like tobacco or alcohol or anything else.

I think you just might be missing an opportunity there, that in the real setting it's used --

MS. HUDSON SCHOLLE: And I guess what I'm saying is that if the health plan could find that has been done and produce that record, it would count in this measure, because it is a ---- the numerator is a hybrid. It allows the plan to produce the record from wherever it takes place.

So the issues, we would just need to make it clear that it's not just the primary care medical record that counts, that they would provide it. So that's a guidance we might be able to provide. Thank you. But thanks for that clarification.

CO-CHAIR BRISS: Yeah. So I just want to say that I think you could -- I think that this is less about this measure today and more about things that you can think about going forward. So I don't actually think it requires much of a back and forth. But Michael, I'll let --

MEMBER TRANGLE: I mean that essentially was my point. It's not an NCQA

issue, because it varies widely according to state law, whether health plans pay for this or not. It's an NQF issue, if we're looking at ourselves as an integrating function between health plans, primary care, behavioral health, hospitals, and this is an area that's not covered by any of them, you know.

It's a CMS issue, almost. But it does get paid for through public dollars, but maybe we haven't quite engaged with the right constituency to talk about it.

CO-CHAIR BRISS: So any other comments on this measure before we vote? Hearing none.

MS. BAL: Okay. Voting is now open.

There you go. Okay. So the vote for overall suitability for 2600 is yes, 19, no, two, and this measure will be moved forward for recommendation. Now we can move forward to 2605.

MS. HUDSON SCHOLLE: Sure, okay.

This is my last one.

(Laughter.)

CO-CHAIR BRISS: We'll do our best to make it particularly painful then.

MS. HUDSON SCHOLLE: Okay. So this measure is really focused on trying to encourage continuity of care for people who are seen in the emergency department, either for an alcohol or drug problem, or for a mental health problem.

So this is not like the previous ones, where we've talked about a population of people with SMI or AOD dependence, right? This is about care, trying to improve continuity between emergency care and outpatient care.

So it's actually modeled off an existing measure that looks at follow-up after hospitalization for mental illness. But in this case what we've done is we've looked at follow-up after an ED visit. So there are actually four different rates in this measure.

Two of them apply to emergency department visits for a mental health diagnosis and for those, the two different rates are ---- they're looking at the proportion of people seen in the emergency department, with a primary diagnosis of mental health, and did those people have a follow-up visit within seven days that also has a diagnosis of mental health, okay. So an outpatient follow-up?

FEMALE PARTICIPANT: Primary diagnosis.

MS. HUDSON SCHOLLE: A primary diagnosis, right. So what we're trying to do is see that if you're seen in the emergency department for a mental health issue, that there's follow-up for that mental health issue. Unlike the follow-up after hospitalization measure, if you're familiar with that one, we don't have a requirement that it has to be with the mental health practitioner. It could be anywhere, as long

as the mental health diagnosis is addressed.

Likewise, we have two rates that are for people with alcohol -- who are seen for an alcohol or drug diagnosis in the emergency department. We're looking for primary diagnosis in the emergency department, and then looking for the follow-up to occur, either within seven days or 30 days on those rates.

So the reason we focused on this is because of the literature that shows the large proportion of people that are seen in the emergency department for mental health or substance use problems, and the lack of follow-up care that has been documented in the literature.

And because we believe that EDUs represents a failure of access to care, and an opportunity to try to intervene and connect people to care. So we've tested this measure. Unlike the other measures we used, this is a claims-based measure only. So we're looking

at claims data. We tested it using the Medicaid analytic extract, which is a claims database that represents a number of states, and allows us to look at patterns across states.

We saw wide variation across states. This Medicaid database is primarily fee-for-service. There are -- the data that's on managed care plan enrollees is less reliable.

In general, we had strong support from stakeholders through our focus group, public comment and our advisory panel, supporting that this measure's important, and encouraging the use of this measure.

CO-CHAIR BRISS: Terrific, thanks. Les, can you walk us through?

MEMBER ZUN: With or without bias?

(Laughter.)

MEMBER ZUN: So okay. So let me make a couple of preliminary comments, and then I'll give my ---- or the Committee's

critique. I should be on. So yes, this is a measure looking at really four rates. So when they present to the emergency department with one of three diagnoses -- primary diagnosis mental health, alcohol or drug dependency, so one of those three, that they get -- not that they get, but they actually follow-up seven days and 30 days after the index visit, which would be the ED visit, to any provider.

Was I clear about any provider, so all those other services aren't part of that, like AA, et cetera, et cetera. So can I now give my bias?

CO-CHAIR BRISS: As long as it's fully disclosed.

MEMBER ZUN: Oh, all right. So here -- as far as I'm concerned, the motivation and desire to get these people in follow-up is meritorious, but this measure does not provide for that for -- we'll say four reasons. So one is they have to have a diagnosis of mental health, alcohol or drug

dependence.

In the emergency department, we do pretty good about mental health diagnoses, but we don't make a diagnosis of alcohol or drug dependency. Now that's a whole another issue about whether we should be doing that, but that's not frequently -- that's not a frequent diagnosis that's given, and we don't go through the screening or diagnostic tools to do that, and maybe one day we will.

So that's issue one. Issue two is resources. Whether they're Medicaid fee-for-service or, you know, they're going to be in a Medicaid HMO, many communities don't have resources for mental health, alcohol and drug dependence, and if they do, it could be months until they get -- until they're able to get into a treatment facility as an outpatient or inpatient or anything.

So that's a little problematic in some communities. So as far as I'm concerned, it's very hard to hold the emergency

department accountable for resources that may or may not be available in that community, and we know there are many communities that don't even have any psychiatric services, let alone substance abuse and alcohol services. So that was the second.

The third is I believe this is displaced responsibility. The responsibility should be on the health plan, should be on the HMO, should be on Medicaid services to get their folks into treatment, not the emergency -- the emergency department has nothing to do with connecting -- you know, we just make the referral, and sometimes we don't even make the right referral because sometimes we don't know what their coverage is, which kind of leads to my next problem is, you know, some say they're on Medicaid, some aren't on Medicaid.

We don't know. We don't verify it. The EMTALA requires us to see every patient in the emergency department and do at least a medical screening exam. It doesn't

say that we need to confirm their insurance status and ensure that they're Medicaid covered, and if they're Medicaid covered, that we provided these services.

So now we have a differential of care between Medicaid-covered patients and other patients. So that's the fourth problem, and then the fifth problem I had was -- I couldn't find in any of the stakeholders or in any of the people who developed the tool were emergency physicians or emergency medicine specialty organizations were represented.

Then the sixth -- is that too many? Can I give a sixth one?

CO-CHAIR BRISS: You can have a sixth.

MEMBER ZUN: If we look at the validity testing, they state, while empiric testing did not support our hypothesis, stakeholders generally supported the face validity. Well -- but if you look who the stakeholders are, the stakeholders weren't the

folks in the emergency department. They were -- I don't, you know, they were outside the emergency department.

So I'm a little concerned about even their own statement about validity. So I would -- so my bottom line is I would suggest that they rework this and look at referrals from the emergency department for Medicaid recipients, and not, per se, that the patient actually followed up in the prescribed number of days. Thank you.

CO-CHAIR BRISS: Thank you. Yes. So thank you for ---- to the Senator from Emergency Medicine.

(Laughter.)

CO-CHAIR BRISS: Do you want to --

MS. HUDSON SCHOLLE: I would just clarify ---- a couple of clarifications. This is a plan level measure. So it is not holding the emergency department accountable. It's actually holding the health plan, or the state, accountable for this connection

happening.

Now the health plans don't like having this accountability either because they say, how are we going to know somebody? But there are ways that they can make that possible, to make that connection happen, and if there's a lack of services available then that's part of their responsibility.

This is a measure that is based on claims-only, and the benefit -- the plan would be responsible for the benefit, for the mental health benefit. So they're responsible for making sure that their services are available ---- that they have a network of services available.

And when we get to the validity, I can ---- we presented the validity data. You want to hold that one? Yeah.

CO-CHAIR BRISS: So maybe since Harold saved me for last when we last commented all open -- so I'm going to take off my Chair hat for a second -- and so, as has

been said by others this morning already, I think you have to take your own temperature on how you feel about the leading a horse to water and encouraging him to drink issue.

I agree. As has been said previously today, I think we're past the point where, you know ---- there was an old Tom Lehrer song where he quoted Wernher von Braun, or he allegedly quoted Wernher von Braun about -- I shoot missiles up, who knows where they come down? That's not my department, said Wernher von Braun.

So I actually think we're past the Wernher von Braun stage of the world, and we're at least partly responsible for whether people get what they're supposed to get and whether they get better as a result.

I was also going to comment on that this is a plan level measure. It's not really holding the ER specifically accountable. So I'll just go around, Mady.

MEMBER CHALK: I didn't notice

what the data were on diagnoses of alcohol in the ER because we know specifically that, in at least half of the states, law still exists to create a disincentive for emergency rooms -- emergency departments to diagnose alcohol because they can't get paid for any resulting conditions that are related to the alcohol.

So that being the case, and it is the case, I wonder how that affects all the rest of the data you're going to collect, or how this measure's going to perform, given that in half of the states and maybe in every state for other reasons, they're not going to diagnose alcohol and drug dependence. Nice to include it, but --

CO-CHAIR BRISS: And I think we know that for many measures, including this one, that the problem -- whatever problem you're trying to solve is going to be -- is going to be under-ascertained to a greater or lesser degree. So it's likely that there's going to be some under-ascertainment of mental

health issues and likely a larger under-ascertainment of alcohol and drug problems.

That doesn't mean that the focus of the measure, which is coordination of care for the people that you do know about, is a bad thing to do.

MS. HUDSON SCHOLLE: And just to --- maybe we're really interested to understand this phenomenon that you've described. We actually found people with these diagnoses, and if you look at the data in the testing report, it's in there.

So we do see variations across the states, and it's smaller generally than the mental illness -- the people with the mental illness diagnosis. But it's present and in the Medicaid data that we looked at, and this was not something we heard from others.

MS. LIU: The AOD rate would also require the chemical dependence benefit. So if plan's responsible for the benefit, then they should pay for the services.

CO-CHAIR BRISS: Raquel.

MEMBER MAZON JEFFERS: I had, I think, two questions about two exclusions and a question about an inclusion. So pertinent to the conversation that Dodi started earlier, I actually think in this particular situation, it's even more relevant.

So someone being referred -- so excluded from the numerator are individuals who have been transferred to subacute residential treatment, and for an individual with an alcohol use disorder that appeared in -- that showed up in the ED, in many cases the most appropriate referral for them is to a subacute residential detox program in the community.

So I'm not really sure why they're excluded from the numerator, and it does explicitly say that only outpatient and intensive outpatient seven day visits are possible. It's actually also true for SMI. Very often the first place you want to

transfer someone who showed up in your emergency department is a short-term treatment facility for SMI, if they are in fact rising to that level of care.

So that was one question about an exclusion from the numerator. Also the measure excludes from the denominator anybody with a secondary or tertiary diagnosis of substance use or mental illness.

So very, very often, people appearing in the emergency room show up because of a trauma injury, and the first diagnosis is related directly to the trauma and the second diagnosis is related to their behavioral health condition.

So very frequently the emergency department coding is in the secondary and tertiary diagnosis for the behavioral health condition. The third question I had is about an inclusion. So included as appropriate care in the community, seven days out and 30 days out is a referral to targeted -- is actually

an engagement with targeted case management.

As far as I understand, targeted case management is a linkage service. It is explicitly not, by Medicaid standards, a treatment service. So I'm not really sure why that is counted as a referral to treatment, because it's not -- technically, it's not actually a treatment service.

CO-CHAIR BRISS: So let's hold the response to that a second, because most of that is sort of about measure specs and reliability and validity, and I'd like us to finish on importance to measure first, if that's okay.

MEMBER GOLDSTEIN GRUMET: So one question that maybe the ED docs can answer, if somebody comes in for suicide or self harm, and it's just thoughts, are they going to get coded if there's no diagnosis necessarily?

I mean I don't -- if not, one of the concerns I would have is that somebody who comes in on suicide, it's not a lesson. It's

impulsive. It's a relationship issue. It's a middle-aged man and had some change in his life -- that potentially some of these cases would not have a mental health diagnosis and need to be followed up.

Second, when people with suicide are referred, we know that a really strong linkage is a crisis line. There's been some good research that shows that that reduces readmission rates, and oftentimes crisis lines can't -- they don't all bill for those services. Some do and some don't.

But in many cases, it can be a really viable source, to keep the person out of the hospital and to keep them safe. And so I just wonder as well, if that's an additional resource that people are using in some cases and it's not going to show up necessarily in billing.

It doesn't mean that the hospital didn't necessarily take on the burden of making what is a good linkage.

CO-CHAIR BRISS: Maybe still another ---- more about specs and reliability and that kind of stuff. So why don't we hold that one too and try to finish on importance to measure and report. Mike.

MEMBER LARDIERI: Yeah. I just had a question. I should have asked it with the other ones too. But with a full range of telemedicine, does that count as a visit. And then, as very shortly we're going to be able to communicate, you know, from my home to my provider, and that will be counted for a CMS visit. So will those things be counted as a visit?

CO-CHAIR BRISS: So the whole series of spec questions. I'm going to try to hold them a little bit longer. Can we try to finish on importance, please? You okay. David.

MEMBER ROBINSON: I got my sign back, so I can ask a question. Okay. So this area's really an area that is a systems

problem. Patients who show up in the ER and who have a mental health diagnosis, when you look at the stats, and there is a HCUP study that was done about four years ago ---- shows that more than 70 percent end up being hospitalized.

By the nature of the ER, the way it's set up, the limitations of what can be done in an ER, and -- I'll just say, the liability of releasing someone with a mental health diagnosis and no setup -- no secured setup or follow-up. So I would think this measure could help in some way ease the concern of the ER docs in that way.

I also see it as a systems issue. So I do believe it's a health plan issue, and so I agree with that from that perspective. I do have some questions about the definition of emergency department.

Since in the mental health world there are many different kinds of crisis intervention-type services, and some of them

are very well advertised in certain states and communities, where patients will show up there as opposed to emergency room.

So my question is whether or not your specifications -- I'm sorry this is a specification question, whether or not your definition of emergency department is that a classic emergency department type, or does it include some of these other types of services ---- could be included?

The other thing about this measure is I see it as exceptionally important, and it's exceptionally important because, as I mentioned before, people are over-hospitalized and put in LOC units, and when you talk to consumers, it's an issue that is high on their list, where they're inadequately serviced in the ER because of the structure, the way it's set up, and are not necessarily referred to these other crisis-type services, which health plans try to make those connections between the classic ER and these emergency services.

And frankly, we could use a lot of help in terms of really pushing that as a mechanism that comes under the ER, as opposed to hospitalization. So this is exceptionally important from a consumer protection advocacy perspective.

CO-CHAIR BRISS: So I'm going to continue to try to keep tabling the spec questions until we get to specs.

MEMBER PATING: All right. So I have several others, but I think it relates to -- I'm concerned with the logic model, which I think is Evidence 1A. So I think in the logic as we want to get ED -- we want to link them, right, to services, whether it's primary care or other things in your other various measures.

This is obviously linking to mental health/substance abuse. Just historically in the last session, we looked at Joint Commission measures, which tried to do this linkage from the hospital, and I think we

found linkage from the acute care setting is just intrinsically difficult.

I will tell you, particularly in parts of California and I imagine in many states that are more rural, I have areas of California where the ED is 120 to 200 miles away from the county where the person is referred.

We have these rural parts of our state which have 1,000 people in them. They have to go -- if they have a psychotic episode, they have an emergency room ride for 100 miles to the ED. They get better and they go back on the emergency room ride, and they don't even live where they came from, because it's rural.

The idea of this linkage, it's a very hard concept, which I guess when I looked at how you're measuring the linkage, I think claims data is just like the worst way to do this. First of all, claims data is really inaccurate when you're looking at what comes

out of the ER and where people are going.

The biggest thing I have about the claims data, particularly when you're looking at these seven days windows, it's too quick in order to do quality improvement. You're giving people seven days from some sort of diagnosis.

There's no chance for a system to like correct itself based on these claims, because they're always retroactive billing, and you're always kind of looking in the rear view mirror.

I would worry that if you're trying to drive a system that wants to build linkage, we're always looking in a rearview window of data from three years ago -- because the HEDIS cycle is one year, based on the data two years ago, and that we're not going to really drive the system.

So that there's a fundamentally kind of logic flaw, and this being the way --- - the indicator, to make these linkage in a

system that is very fragmented. I think you need to design something that's much more real time, with a much more realistic sense of how people get into ERs and out of ERs. So that's just my thought.

CO-CHAIR BRISS: So I think I agree, that that's -- that that has to do with the fundamental logic of the measure. Do you want to comment or respond?

MS. HUDSON SCHOLLE: So I try to think about this like the way Medicare has thought about how they're trying to disincentivize readmissions. It seems to me that a health plan is accountable for making sure that people who -- get access to the services that they need.

If they're responsible for the mental health benefit or the chemical dependency benefit, and they're paying for the ED visit. So ---- and if they want to avoid paying for more ED visits and making sure that their people get into appropriate care.

So then the question in my mind is how does the health plan work with the emergency department that it is paying and the chemical dependency services for the primary care providers that it's paying, to try to make those things happen?

What's happening on the primary care side is by using medical homes, they're using ---- working to help -- health plans can help to encourage hospitals to notify medical homes that their patients are being seen in the emergency department.

Can they do that in a special way for people with mental health and make that a priority and connect -- so that where people are going for the emergency department visit, it happens before they leave the hospital, not based on data that comes to the health plan, you know, a month later after the visit.

So I agree with you. It has to be proactive. But the responsibility of getting somebody from the emergency department, where

alcohol or mental illness is their primary reason for that visit, to getting them into care, that does have to be proactive, prospective, based in the ED, and the health plan's paying for both pieces of it.

So I believe there are ways that they can help their network of providers work together. It's not easy, but that's the tool. Making the emergency department responsible for that doesn't seem reasonable, as Leslie has suggested.

Making the primary care provider who doesn't know, or the mental health provider who doesn't know, responsible also doesn't seem reasonable.

It's the place that's paying for both sets of care that ---- what our measure is kind of arguing, that's where the level of accountability is and that's where the systems for communicating and sharing information should -- the accountability for developing those systems should occur.

MEMBER PATING: And I'm just saying, I don't think this is the measure that's going to drive the system change at the level that you're wanting it, because it's too distant, too late, too remote and it's just not in the logic cycle of what happens in emergency rooms.

CO-CHAIR BRISS: Thank you, although it seems to me that almost any measure for public reporting or pay for performance is going to be -- if you expect it by itself to move the system, it's almost never going to do that, because they're all too late and too infrequent, right?

So what these measures are kind of doing are highlighting important issues, and if you really wanted to do quality improvement, you'd have to be building other systems to sort of complement the measures, right. So Larry?

MEMBER MILLER: I had a quick comment and a quicker question, and Susan

actually addressed my comment about behavioral health homes. I think that's an excellent place for them to deal with emergency rooms, and actually we have a metric in our system that deals with that in terms of the number of folks who are seen within a certain length of time.

My question has to do with why seven and 30 days? I mean if you really want to see them and have them seen, and this is a problem, they should be seen within seven days. You give a system 30 days, they'll take the 30 days. I really think that this is a concern for my part.

MS. HUDSON SCHOLLE: So we had a lot of -- so we basically followed the existing measure for hospitalization, follow-up after hospitalization of seven and 30 days, just to be -- make it easy to program frankly.

But I think there were concerns actually about the seven days, is that too short? The 30 days allows us to see whether

plans are able to make it up. So that's where the logic came from.

MEMBER MILLER: Putting my hat on as a clinician, I think that's just -- and an administrator in the system, I think it's just too long, because what happens, people get readmitted. They might get lost in follow-up. So I think it's, you know, an issue again.

MEMBER ZUN: Just a few more comments. So first of all, I can envision a scenario where the information about the compliance with this measure being out there in the public, and a hospital administrator saying, oh, you guys in the ED only referred ten percent of those with mental health and substance use and -- or ten percent of the patients you saw in your ED actually followed up in seven days and 30 days.

Although you might be well-meaning in having the health plans and the providers be responsible, I've got to tell you that it's going to come back to haunt me as a provider,

because this data gets out in the public domain and the rest is history. So that's number one.

Number two item is there's about 145 psychiatric emergency services in the United States. Is that considered an emergency department visit, or is it not? Some are free-standing, so a psych patient may go to a PES. Some are next to an ED and go to the ED and then they go into the PES. So that's a question.

Third is what do we do about that patient that boards in the ED for a week, waiting for an inpatient bed and finally we decide that they're well enough to send them home? So they've already hit the wall as far as they're supposed to be getting an outpatient visit, but in the meantime, they're waiting.

There was actually a study done in California, where I think it was 20-some percent of the ED directors say they have

psych patients waiting in their EDs for over a week to get an inpatient bed. I'm sure I can give it to -- I'm sure David is familiar with those problems.

Then lastly, I wanted to comment a little bit more about the claims data, because we know that the claims data doesn't always accurately reflect what the patient was seen for in the emergency department. It may be based on what the best billable rate may be or those kind of things.

So the claims data is somewhat suspect for that, and we may actually -- I think some EDs might not put a substance abuse or mental health issue on there because the question of reimbursement or payment ---- like a trauma patient.

If you put down on a trauma patient's chart that the patient has a substance abuse problem, then we don't get paid for their services.

So I'm not sure we can fix the

whole system, but it's a much bigger problem overall than just, you know, being addressed in the measure. Thank you.

CO-CHAIR BRISS: So we're continuing to get a lot of questions about what counts as an emergency department, what counts as a service, those kind of things, and I really want to push those into the next discussion of reliability and validity.

So I think I have -- Julie, is your card still up? So I think we're out of cards for importance to measure and report, and I'd like to try moving us to voting and see how that goes.

MS. BAL: Okay. Voting is now open for evidence for 2605. Now it's open. My mistake.

[VOTING.]

MS. BAL: Okay. So the results for evidence for 2605 is high 9, moderate 9, low 4, insufficient 0, insufficient with exception, 0 and we can move forward to gap?

I'm assuming we can vote now? Yes. Okay. Voting is now open for gap.

[VOTING.]

MS. BAL: Okay. So the results for gap for 2605 is high 17, moderate 5, low 0, insufficient 0. And now we can vote on high reliability. I'm sorry, high priority.

[VOTING.]

MS. BAL: Okay. The results for high priority for 2605 is high 14, moderate 6, low 1, insufficient 1, and now we can move forward to scientific acceptability, as everyone wanted to talk about.

CO-CHAIR BRISS: Clearly, we have lots of questions that have been coming around about the specifications and the reliability. So at a minimum, there are lots of questions that are about what counts as an ER and there's a lot of questions about what counts as a service, and there are lots of questions about how much of all of that stuff can we get out of claims data. So can you open this up

for us?

MS. HUDSON SCHOLLE: Sure, okay. So first, let's talk about the denominator, which is looking for -- what we're looking for is a particular kind of facility with a particular kind of billing code. And so the specifications that -- we used specifications from HEDIS, from how we define what an ED visit is, and so it's the facility plus the code.

I believe that the -- so if the psychiatric facility, if these special psychiatric ED facilities are using those codes, then they would show up in the claims data. There were questions about the exclusions.

So the way this measure works is people that are directly admitted to an inpatient or facility. So this is really for people who are discharged home, right? These are people who aren't discharged to an inpatient or residential setting.

If they are, if that happens, then they get excluded. From the denominator, right. They're not in the measure. This is looking at people that get discharged home.

So those people that are sitting in an ER and then go -- and waiting for that inpatient bed, they'll go -- they're not in this denominator if they go directly to that inpatient setting, if they go outpatient.

We start the clock at the point where they go outpatient. So this isn't going to deal with those problems.

I'm sorry, the inpatient? If they go -- many detox facilities are subacute residential detox. It would be a discharge from the hospital ER into the subacute service. That would be included as follow-up, then, okay. So if they -- the plan has to be responsible for the chemical dependency visit. So as long as the detox was in there, that would count as a -- that would count to the numerator, right?

So if they're discharged home and then they go to detox, then that would count.

CO-CHAIR BRISS: I think the question is if they're discharged straight to detox.

MEMBER MAZON JEFFERS: I think there's some confusion.

MS. HUDSON SCHOLLE: If it's an inpatient or residential setting, then they're out of it, because this is only people that are discharged to the community.

MEMBER ROBINSON: Right. This is a system measurement. So you're saying if the system from the ER gets the person to the residential, which they're going to get there by transportation that is guarded, then it doesn't count here.

What you're really looking at are the people who actually go home and have to show up some place after that.

MS. HUDSON SCHOLLE: That's right. Thank you, Rhonda.

MEMBER ROBINSON: So an ambulatory detox would work, but not an inpatient-based detox or subacute detox.

CO-CHAIR BRISS: Do you have a follow-up on this point?

MEMBER MAZON JEFFERS: I think the numerator clearly says -- I'm sorry. I think the numerator clearly says that the visit has to be outpatient, an outpatient visit, intensive outpatient encounter or partial hospitalization. That's what it says.

So then it would mean that these other community providers that we were talking about earlier, that are part of the ACM continuum of care for people with substance use disorders that are community. It wouldn't be a direct transfer. It's not considered a hospitalization.

There are subacute residential treatment programs that are community providers, that may not be paid for by a health plan or by Medicaid, but are still very

appropriate next steps for the individual who left the ED to be referred to.

MEMBER SHEA: So this is for -- this is a health plan measure. So if those services are paid for by the health plan, then the health plan's going to capture them? Is that what you're saying? It's not now.

MEMBER ROBINSON: No. I think the question is when you're talking about community-based services, let's talk about like three quarter houses. Is that what you're talking about?

MEMBER MAZON JEFFERS: Any residential substance abuse treatment program.

MEMBER ROBINSON: Residential implies that the person lives there at night. There's a room and board, and that would be excluded from this measure.

MEMBER MAZON JEFFERS: Exactly.

MEMBER ROBINSON: What she's trying to measure is the strength of the system, to get the person to come from home

and to follow-up after an ED visit. Going to a residential, they're living there. They are room and board there. That's the issue.

MEMBER MAZON JEFFERS: So nonetheless, it's an -- so the person would come from home. They would leave the ED. They would go home, and then they would go to a residential treatment facility, depending upon their level of severity of their substance use disorder, and it would be a very appropriate step in the continuum of care for them.

So rather than qualifying to be in intensive outpatient care, the person's substance abuse disorder might be severe enough that they require a more controlled recovery environment.

MEMBER MARK: And I think the specification issue is that a lot of times those residential will be carried as inpatient. So if you're only counting outpatient as follow-up, you wouldn't be

capturing them. So can you just clarify if inpatient follow-up is captured?

MS. HUDSON SCHOLLE: Actually, and people that get admitted within that time frame, within that next 30 days. So if they got admitted, they would also be excluded, because we're looking for the outpatient care for people that remained in the outpatient setting.

So I think they're either -- I believe that the description that you have, I'm pretty -- I know that it is -- it's an exclusion, because we think if they're connected and they get into that inpatient or that residential setting, even if it's not direct transfer but it's transfer within this time frame, this time window, then they're excluded from the denominator.

We're really trying to focus on people that go to the community and are expected to get care in the community. Okay, and so let me just continue down some of the

other questions that were raised.

Would crisis services count? Yes. Mobile unit services are in our codes. Telemedicine.

We are working on the telemedicine code. CMS has just introduced a series of telemedicine codes that don't -- are not included here, but it's one of our projects at NCQA to actually update all of our codings, all of our measures that require visits to include telemedicine. So we're working on that. It's not represented yet.

We focused on primary -- so for the numerator, okay, and the denominator. For the denominator, we're looking for people with a primary diagnosis. We had a lot of discussion with our stakeholder groups and our advisors about should we include a diagnosis of mental illness or alcohol or drug dependence that occurs anywhere on the claim, or should we look for primary?

Well, that actually got into the

problem of how quickly should you require? So if it's a primary diagnosis that is putting them in the emergency room, then requiring that there be a follow-up within seven days or within 30 days with that same diagnosis or that, you know, mental health diagnosis if it was a mental health visit, that seems like a reasonable requirement for the health plan, to make sure that that visit happens.

There were concerns that if we said a secondary diagnosis, that you wouldn't know whether that was really precipitating the event and that they weren't in care. So that's why -- that's why we stuck with primary. We felt like if you said primary, then it was a lot easier to say and you have to have a visit within seven days or within 30 days.

So that's why we've combined those. Others thought that it would -- requiring such follow-up in that time frame might not be warranted.

So and in terms of suicide, this is based on diagnosis. So if they put suicide, then they'd have to use a mental illness diagnosis for this to count, okay.

CO-CHAIR BRISS: Let's go to David and then David.

Let's see if we can finish reliability first, I think, because clearly there are some complicated issues being raised, and I think that it's hard enough to deal with one issue at a time. So David.

MEMBER EINZIG: So I think I'm putting this in the right place with the reliability.

(Laughter.)

MEMBER EINZIG: I'm having a lot of trouble with this concept of making it happen, making the patients go to their follow-up visits. So I like analogies. I'm thinking of, you know, making my kid eat spinach, right, because spinach is good for the kid.

MEMBER PATING: It's seven days. I doubt that he'd eat it.

MEMBER EINZIG: Yeah, right. Like that's going to -- Mike, you don't know my kid.

CO-CHAIR BRISS: I think admitting parenting limitations is a slippery slope that I don't want to start down.

(Laughter.)

MEMBER EINZIG: But so the horse and cart concept maybe is backwards. I mean maybe the system should be the horse and the patient is the water. I mean if we want to make this happen, then let's go to the patient's home, right? I mean otherwise gas in a cop car or a limousine or something, and transport the patient.

It's just that I have trouble using this as a quality measure, when control in this sense I think is a myth. You can influence behavior, but I don't think that you can control the behavior.

CO-CHAIR BRISS: Mady.

MEMBER CHALK: Question. Am I correct that you said that this is only in reference to patients who are going to be followed in the community, so that the most seriously ill patients are excluded from this, who need some kind of other level of care are excluded?

MS. HUDSON SCHOLLE: So basically they get excluded because they get the care, right? So the ones that are sick enough -- okay. If they get -- okay, so let me -- so if people are seen in the ED, and they either are transferred to an inpatient or residential setting at that time, through a direct transfer, or within the 30 day follow-up period, right, then they've gotten transferred to intensive care, right?

If they are not, okay, so they get excluded. So what we're trying to look at are the people that don't get to that -- that don't have that intensive need and get into

that intensive setting quickly.

We're looking at the people that are remaining, the ones that appear to be discharged to community and not getting follow-up, and we're looking to see are they getting follow-up, but they're not going to the intensive setting.

MS. LIU: Yes. Just to add to that, you know, the exclusion is about patient who were directly transferred into the inpatient or residential setting right after the hospitalization, or they got readmitted, which then started a follow-up period into a residential or inpatient setting.

So they were getting care there. Therefore, we exclude them from this measure and health plans are not penalized for those patients, because those are excluded from the denominator. So that's the focus of this measure, is who are discharged in the community and need follow-up care.

CO-CHAIR BRISS: So Raquel.

MEMBER MAZON JEFFERS: So interestingly, I think that the individuals with the primary substance abuse and mental health diagnosis that appear in the ER, the sicker population that are more likely than to be excluded.

So it's the people who show up in the ED with a secondary and tertiary mental health and substance use diagnosis, who are probably more appropriate to be referred for the outpatient service.

So I'm just concerned you're going to miss a lot of people by excluding them, because you're only looking at the primary diagnosis, and then oh my gosh, there was one other thing.

Oh. I did want to share that there are system solutions that have been very effective, that I've actually seen in New Jersey, where there are accountable care organizations that are doing intensive case management like behavioral health homes, of

managing high utilizers.

So that that care coordination team receives a message the second one of their clients hits the ED, and they are greeting that person in the ED and able to link them to needed services.

So I actually do think that this measure, in terms of the horse and the water and the cart, is a really nice diagnostic of the system's ability to devise and invest in a response for these very complex patients.

I do think it's possible. I've seen healthy systems or systems that are trying to address this be successful.

CO-CHAIR BRISS: Rhonda.

MEMBER ROBINSON: I just want to answer you, Raquel. I think you're absolutely right. There are systems that work when they have that connection with the ER, and that's really the critical part of this. So the ER has to notify that system. Otherwise, that connection doesn't happen.

So I think there's two things. One, this measure does belong with the health plan, but there needs to be a counter-accountability to the ER at a facility also, because they have to do that.

The other thing, in terms of whether or not this touches seriously ill patients. I think the distinction is do not assume that people who are hospitalized are the only ones -- are the ones who are seriously ill.

There are seriously ill individuals who don't require hospitalization, but do require acute ambulatory care. I ran a system for 18 years, myself as a capitated provider, and the majority of the patients did not require inpatient hospitalization.

They really required connection to treatment and feeling connected to treatment. And so, you know, there were a lot of things that we did in order to do that, to cut the hospitalization rate and improve the

satisfaction.

So I don't want you to think that this is cutting out those who are ill. It's not. It's just that the inpatient is not or the residential is not the only solution for these individuals, and what this is trying to do is to bring accountability for those who don't need that, and not encourage the ERs to admit these patients as the only solution for these patients, but to use other types of services instead.

CO-CHAIR BRISS: Thank you. So I'd really like to get -- to get very quickly through anything else about the measure specs. So Les.

MEMBER ZUN: I only have two more comments. So what happens when I see someone that's an alcoholic, and I refer them to our referral source, which is Garden State Mission, which provides -- it's basically a shelter, and it provides good services for those that are alcohol-dependent. So would

that count? So that's my first question.

Then my second is kind of a comment, because I think this measure only hits the tip of the iceberg, because we did two studies of undiagnosed mental illness of patients presenting to emergency departments. In the adult population, you won't guess what percentage have undiagnosed mental illness.

We used a screening tool for 16 diagnoses. Fifty percent of the inner city patients that we saw had an undiagnosed mental illness. Second is so we did that on kids, because there was a kid tool that we used, and 45 percent of the kids who come in with colds and flus and aches and pains and strains and nothing psych-related had screened in for a psychiatric diagnosis.

To me, if we're trying to get at it, there's a better way of getting to those that need help. Just because we put that -- you know they have to be really sick to get that mental health diagnosis.

But someone should be responsible for screening these people and getting them the therapy, and if the health plans want to put mental health workers in the EDs, I'm all for it.

CO-CHAIR BRISS: So Mike, I hope this can be the last word on the spec stuff.

MEMBER TRANGLE: Well, in some sense I feel it is a little bit of a summary in my own mind. It feels like I'm hearing both -- actually, it feels like I'm hearing 12 sides of the argument, not just two sides of the argument.

CO-CHAIR BRISS: I think the right number is more like 22.

MEMBER TRANGLE: Whatever the number is, but and I find myself agreeing with the majority of them or at least a plurality, you know.

If I try and make sense of what do I think and what makes sense action-wise, it really is like clearly this is sort of an

early stage of a tool, where definitions are imperfect, and the world is structured in a way where in terms of cultural issues, the flow of ED, what's paid for by health plans, you know, as well as the specs, mean that we're not going to be capturing everybody.

On the other hand, I totally agree with Raquel, and we have an in-service social worker who works with these high utilizers and spends a lot of time getting them connected so they get insurance, you know, and getting them connected so they can then start getting the treatment and other kinds of things.

I can see the utility and I can really see how it does a lot of good. This doesn't seem like I've heard anything about competing measures here. It's not like this is an area people are vying; I want to be a vendor. It's going to be a high profit margin, you know. So part of me is basically saying yes, this is imperfect.

I could focus and perseverate on

imperfections, you know, or I could say God, this is a new effort in an area where nothing's being done, and we all agree that it's crucial to try and get started on this, and then hopefully see how it works and we improve the definitions and start partnering with the states and CMS kinds of stuff for the psychosocial elements that aren't included by health plans.

So I guess I'm saying -- I guess I'm giving you a summary and saying it's probably worth voting for, even though I have trepidation, because it's such an --

Well you could argue about that, whether you know, how it can improve without stopping it here.

CO-CHAIR BRISS: So on the reliability, on the spec side, I keep trying to close the discussion and cards keep coming up. And so what we have here may be a failure to communicate. So if there are things about the specs that haven't already been said,

Raquel and then David please. Is your card still intended to be up?

MEMBER MAZON JEFFERS: The only issue -- are we on validity?

CO-CHAIR BRISS: No. It's still reliability of the specs, and only if it's not already been said. Okay. So reliability testing we haven't really talked about. So Les, can you very quickly walk us through the reliability testing.

MEMBER ZUN: Well, the reliability testing was a little problematic, in that it was looking at inpatient follow-up and even in their application, it said that the -- let me see if I can pull this up -- the validity was not there. Let's see. Let me get back to this.

Empiric testing. We did construct validity by exploring whether states perform in this measure, related to the rate of inpatient hospitalization for mental illness. I'm not sure how we're going to --

CO-CHAIR BRISS: We're still trying to talk about reliability here.

MEMBER ZUN: I'm sorry. I'll get there. Sorry.

MEMBER PATING: I didn't really understand that. I've looked at this now four times, and I've tried to understand the reliability, how you did it. You had -- what you were comparing apples to apples, what were the apples, and it was a chart review or it wasn't clear to me.

You compared one claims set to another claims set. That's what it looked like.

MS. HUDSON SCHOLLE: Okay. There are different ways to assess reliability for different kinds of measures. This is a claims-based measure, and so for claims-based measures, the approach that we used to look at reliability is something that's -- it's called a signal to noise reliability metric.

So the idea of this metric is to

say can you -- can you pick out one health plan from all the other plans? Is there really a signal here? Does this measure -- or in this case we used state level data. So can you really identify a result for an individual organization or entity that you're trying to characterize?

So the statistic, the signal to noise statistic is based on a beta binomial model, and the numbers are actually shown right here on the screen. Basically what influences this measure is how much variation is there across the different units of analysis, across the states, and what's your sample size.

When you have a good sample size and a lot of variation across states, as we've demonstrated, then you're going to have a highly reliable measure. What that's going to tell you is that it's not going to change -- you can really say, you know, this plan is different from the average of all plans, or

this state is different from other states.

You have a lot of confidence that when you pull this, when you draw another sample, when you calculate this again, you're going to get the same result, because you've got a big denominator, and you've got a lot of variation across the organizations or entities that you're measuring.

So these measures are highly reliable, okay. This measure is highly reliable based on our testing, because this statistic ranges from 0 to 100, and we're presenting reliability results of .9.

CO-CHAIR BRISS: Questions or comments about the reliability of the measure. Yes.

MEMBER ZUN: I understand the methodology you used. I'm not familiar with that methodology, because if you look at the definition, reliability describes how one can confidently distinguish the performance of one physician from other.

I've got to tell you that in my face, the face reliability to this just doesn't -- it doesn't make sense to me. I have two emergency docs and they rarely come up with the same mental health diagnosis.

MS. HUDSON SCHOLLE: This is about the reliability of the measure, of the performance of this measure. When you do it over and over again, and yes, and I apologize if it says something about -- it says of one physician or an accountable entity. In this case, we're looking at the entity, not a particular physician.

This is looking to understand when you implement this measure, can you really pick out one entity from everybody else? I mean does it give you a signal? That's what we're trying to get at with this statistic. It's influenced by how much variation you have and what your sample size is.

Any claims-based measure is going to have a pretty good -- generally, they have

pretty good sample sizes, and this measure has a lot of variability. So that's why you see really high reliability results from this statistic.

DR. BURSTIN: This is a pretty standard approach for almost all the claims-based measures we've got, in terms of the signal to noise reliability analysis.

CO-CHAIR BRISS: But it might help. Can somebody say in simple words what a .9 means in this context? What does the number really mean?

MEMBER SUSMAN: High reliability.

(Laughter.)

MS. HUDSON SCHOLLE: The statistic ranges from 0 to 100, right, or 0 to 1, and so we're getting a .99. That means really reliable.

MEMBER ZUN: Highly reliable.

MEMBER SUSMAN: The chances are relatively small that this is due to random variation. There's probably a true

reliability here as opposed to -- you can pick out differences meaningfully on a consistent basis.

MEMBER ZUN: Maybe I'm confused about it.

MEMBER SUSMAN: I think you're confusing maybe validity and reliability here.

MEMBER ZUN: But if you read the ranges, percentile across states from 10 to 90 percent, and they're all in the 10 to 90 percentile, I'm a little -- that's like the whole universe of --

I mean of course you're going to have a high reliability if your universe is from 0 to 100. It says right there 10 to 90th percentile across states is very, very high. I'm a little --

MS. HUDSON SCHOLLE: This is not the performance rate that is -- so the way -- it's just the way this measure is calculated, and this is how you summarize it and the statistician who does this is not here to

explain it better.

CO-CHAIR PINCUS: I see you're cutting off at the outlier.

MS. HUDSON SCHOLLE: Yeah. But basically what it shows you is that we have very good reliability in the different places where we've looked at it in the different states, because you've calculated for each state, and then you summarize it.

If you're asking about how much variation that we have in performance, we see a lot of variation in performance, and that's presented in a different place in this report, where we do see that the performance rates range across states, with an average of -- and that's in Table 8 of this document, where we see that the median for the mental health seven day follow-up is 74 percent. It ranges from 42 percent to 80 -- to 90 percent.

It's under Meaningful Differences, okay. All right.

CO-CHAIR BRISS: So we've had what

I think is actually a fairly long reliability conversation. It's clear that people have different points of view about the specs, and maybe about the usefulness of the statistics that have been presented. But does anybody else have anything that they want to raise that hasn't already been raised?

About reliability, specifically about reliability. And if not, I think I'd like to try to move us to a vote on reliability in particular.

MS. BAL: Okay. Voting is now open on reliability.

[VOTING.]

MS. BAL: So we're all at 19. Did everybody get a chance to vote, and did anybody step out that I don't see perhaps?

There we go, thank you. So for 2605 reliability, we have high 15, moderate 5, low 0, insufficient 2, and we'll move forward to validity.

CO-CHAIR BRISS: So Les can you

tee it up for us on validity?

CO-CHAIR BRISS: Yeah. We've already talked about some of this, I think.

MEMBER ZUN: So I think there's been a lot of questions and concerns about the validity of the data, both the numerator and the denominator, and in the interest of time, I don't think I'll go much more into it.

CO-CHAIR BRISS: So there are lots of cards up, and maybe we'll start with Raquel and come this way this time.

MEMBER MAZON JEFFERS: So just very quickly, the measures were only tested on Medicaid plans using fee for service claims. So I was just concerned about the applicability across the board.

CO-CHAIR BRISS: So Bob.

MEMBER ATKINS: I have two concerns about validity. One is, as I understand it, it's a claim-based measure, and across markets, across states, across regions within states, a substantial part of the

outpatient services, counseling, psychotherapies whatever, are not provided by Medicaid providers, and it's totally legitimate. It's the right thing.

So people are getting what they need, and it's my understanding that we won't be able to find that with this methodology. So I question about the validity of that. The other is the validity of applying this to the health plans.

This is a system measure, it seems to me. The health plan, the state, the provider communities, plural, are all part of the solution. I'm concerned about making this the accountability of health plans specifically.

I think it's a really important issue, and there's a way to apply it to the systems of care or the lack thereof. I can go on and on about this, but I'll stop right there.

MEMBER PATING: Section 2b2.4 on

the validity. So my understanding of how you tested the validity was that you had a hypothesis, that if you get people connected in these states that had various reliable rates, that you would somehow be able to show that they had less hospitalization.

But your own conclusions were that there wasn't significantly less hospitalization, and that the testing didn't support your ultimate outcome finding. It was because of the effect size was sort of too small, I guess, or there wasn't too much -- even though there's this difference between states, this difference between the hospitalization rates wasn't significant enough so.

Then you went to -- then you backed it up by saying well, it's still face valid, and I think that we're hearing concerns around that methodology. So can you just speak to that gap?

MS. HUDSON SCHOLLE: So we could

have just repressed the data on our validity testing. So this is -- it's typical that when we do measure testing, that we would try to look for correlations among measures and patterns across states, and try to understand, you know, how does this measure relate to something that we've hypothesized that it might be related to, and it doesn't always work out, and this is what we found.

So what -- part of the reason why we think we couldn't find that difference is that we've not actually varied, you know, a correlation.

If you're looking for a correlation, with something that doesn't vary very much, that has a narrow distribution. So one thing that's narrowly distributed, and you're trying to find a correlation with something that's not.

If it's not, then you can't find a meaningful correlation. So that's -- that was our explanation for our results here, that we

presented in Table 3 and Table 4. We showed these results to others, and they said well, we're not ready to throw away the measure based on these validity results, okay.

That's basically what we're telling you is. When we showed this to stakeholders, we explained our problem. They said we can't explain the results, but we still think the measure has merit, and that's what the base validity is about. It's saying the measure has merit without this hypothesized relationship.

I did want to respond to one other validity concern that was raised, about who's responsible and are claims data going to pick up the data, pick up these services. The way that the specifications are set up, we require that the plans report --

That the plans be responsible for the members' chemical dependency benefit or mental health or behavioral health benefit, as well as the general medical benefit, so that

we're not in that situation of its care that's not in the benefit. If it's in the benefit, then the plan is responsible for that individual.

CO-CHAIR BRISS: So there are more and more cards up, sort of four more cards up on sort of any additional comments on validity that haven't already been made as quickly as possible, starting with Vanita, please.

MEMBER PINDOLIA: So my question on the validity is I noticed that during the stakeholder public comment, it was addressed of concerns of health plans having a lag time, the ER claim coming to enact upon a seven day follow-up. So when you were doing your validity search, did you notice what kind of gap there was from when a claim gets to a health plan?

The reason I'm bringing that up is I think it helps address to what Bob has said, and I think others have said. This really is more than just a health plan. This is really

needing the providers also to be held accountable for this joint effort.

MS. HUDSON SCHOLLE: So we couldn't analyze that in claims, but -- in our analyses. But it would require a proactive system, as others have talked about, of the ED contacting them and the health plan mediating that in some way.

CO-CHAIR BRISS: Yeah. I think we've actually raised this issue already, that this is a retrospective, you know, sort of reporting measure. It by itself isn't, you know, wouldn't be immediately useful as an internal quality improvement tool. Tami.

MEMBER MARK: Yeah. I've looked at the validity of post-discharge follow-up from psychiatric and substance use hospitalizations, which is very parallel, and it's actually hard to demonstrate that it has an effect in observational data, because you have such a selection bias.

But if you look at some of the

studies that have been done, you know, more rigorously in terms of randomized trials, you know, you do see that there's evidence that post-discharge follow-up does reduce readmissions, and I would think that, you know, you can use that evidence since this is, you know, a similar type of thing, to support the validity of the post-discharge follow-up after ED. But I don't think in your application you cited any of that research.

MEMBER SUSMAN: Thank you. I don't find that this "failed" association/correlation with regard to your validity testing is particularly telling or concerning. I mean I can think of lots of different confounders in this.

So I think as a group, as we look at this, I'd say yeah, well that didn't work out. But it doesn't affect me in thinking about the validity testing overall.

I would hope that in the future, we would have other approaches to validity

testing, and also look at some of the issues that Les and others have raised, when you get down to okay, does this really reflect appropriate, accurate data as we're going forward.

CO-CHAIR BRISS: So thank you, and I'd like to suggest that we try to vote on validity please.

MS. BAL: Okay. Voting is now open.

[VOTING.]

MS. BAL: Just waiting for two more votes. If everyone could please vote, thank you. Okay. The results for validity for 2605 is high 3, moderate 9, low 8, insufficient 1. So this is actually in the gray zone, yes, and we can move forward to feasibility.

MEMBER ZUN: I think it would be best that I not directly comment on this, because I'm afraid I might bias someone in the room. So with that, I'll close.

CO-CHAIR BRISS: So to quote my next door neighbor, it's claims data. So it appears feasible on its face. Any additional thoughts that need to be raised?

(No response.)

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

MS. BAL: Okay. The vote is now open for feasibility.

[VOTING.]

CO-CHAIR BRISS: I told you we'd be hard on you for a going away present.

MS. BAL: Can we get one more vote please. If everybody could make sure that they voted and point at me please?

Okay. So the results for feasibility for 2605 are high 5, moderate 13, low 2, insufficient 1, and we can move forward with usability and use.

CO-CHAIR BRISS: And it seems to me that we've had a lot of discussion that sort of touches on the usability of this

measure. So are there additional comments that relate to usability that haven't already been addressed?

(No response.)

CO-CHAIR BRISS: Hearing none, let's try to move to voting please.

MS. BAL: Okay. Voting is now open for usability and use.

[VOTING.]

MS. BAL: So we're only at 17 votes. If everybody could make sure to vote.

(Pause.)

MS. BAL: So the results are for usability and use for 2605 is high 5, moderate 8, low 5, insufficient 3 and okay. Consensus is reached on that one, and we can move to overall, unless there's --

CO-CHAIR BRISS: It's over the 60 percent line. So any closing arguments before overall suitability? Yes.

MEMBER ROBINSON: I know that we've looked at a lot of measures and people

are tired and all of this, but I just want to emphasize that this is probably one of the more important measures that can help improve the system where there are significant gaps.

Again, I want to emphasize the gaps cause harm to consumers with unnecessary hospitalizations, because of the way that ERs function and have to protect themselves from the liability of releasing someone without a clear follow-up. So I just really want you to keep that in mind as it relates to this.

It's not perfect, but it's one that really addresses a huge gap, and one that has a potential for helping consumers in the long run.

CO-CHAIR BRISS: And I suspect we may get a counterpoint. Les.

MEMBER ZUN: Actually, I want to agree, that I think we're on the same page. I think we want to ensure that providers are making sure that we capture these patients in care. I'm afraid that this measure does not

-- is not suitable, does not meet that basis.

I think I would send it back to NCQA and ask them to revise it with some more input, to really look at what -- to really measure and look at what behaviors they're trying to change in providers. So I think the intent's there; the delivery's not.

CO-CHAIR BRISS: So Mike, I'll give you the last word, and I think we know that there's a diversity of opinion on this measure already. So quickly please.

MEMBER LARDIERI: Yeah. I sort of wrap my head around this as bring it down from the health plan. I'm going at risk. As a health system, I want to do this. I want to do this in real time.

So I think it's important, because we're changing. It's going to be a whole different payor. It's not going to be the health plan. It's going to be me. So I really want to look at this really closely.

CO-CHAIR BRISS: So let's tee up a

vote please.

MS. BAL: Okay. Voting is open for overall suitability.

[VOTING.]

MS. BAL: Okay. The results for overall suitability for endorsement for 2605 is yes 16, no 6, and this measure will be moved forward for endorsement.

CO-CHAIR BRISS: So I'd like to try -- we're a bit behind. I'm still sensitive to the fact that many of us are going to be leaving for planes. I think we'll be okay, because the four remaining measures are related to each other, and I'm hopeful that we can pick up some time on the last four measures.

I'd like to try to set up -- to grab lunch and bring it back to your desk in the next ten minutes or so and --

NQF Member and Public Comment

CO-CHAIR BRISS: Oh, I'm sorry, and we also need to open the lines for public

comment please.

OPERATOR: Okay. If some of you would like to make a comment, please press star and the number 1.

(No response.)

OPERATOR: At this time, there are no public comments.

CO-CHAIR BRISS: It appears there are no comments in the room, and so ten minutes to set up lunch, and then let's restart.

(Whereupon, the above-entitled matter went off the record at 12:38 p.m. and resumed at 12:55 p.m.)

A F T E R N O O N S E S S I O N

(12:55 p.m.)

CO-CHAIR PINCUS: So we're going to get started now. So, boy, I'm glad Peter was able to really do an exceptional job in getting through the stuff in the morning.

I had to keep my mouth shut, which is always very hard for me. I had to keep my mouth shut because it was hard for me, but this afternoon we're making a change in the schedule, so that we're going to deal with the PHQ-9 measure first, and then deal together with the 6-month and 12-month one. It means that we're going to have to like vote twice each time we do a vote for the second one.

But we can do that, right? So, you know we vote twice for importance on the six-month one, importance on the 12-month one, you know, that kind of thing.

MS. DORIAN: We might not be able to, for the voting itself --- we can discuss them both at the same time but then for the

voting we have to go through everything because of our software. Everything for one and then everything for another. Depression/Major Depressive Disorder

CO-CHAIR PINCUS: Okay, good. So is the measure developer here?

MS. PITZEN: Yes, thank you.

CO-CHAIR PINCUS: Okay, good. So can you tee up the depression utilization of the PHQ-9, the 0712 measure for us?

MS. PITZEN: Sure. This is Collette Pitzen at Minnesota Community Measurement. Can you hear me okay?

CO-CHAIR PINCUS: Louder.

MS. PITZEN: Okay. Is this better?

CO-CHAIR PINCUS: Yes.

MS. PITZEN: Okay. The measure that we're talking about first today is depression utilization of the PHQ-9 tool. This is a paired process measure that seeks to promote frequent use of the PHQ-9 with

patients with major depression or dysthymia, adult patients aged 18 and older, and also supports the outcome measures of depression response and remission.

So this measure is capturing all patients who are seen within a four-month period, who have major depression or dysthymia, and have a PHQ-9 tool administered at least once during that four month measurement period.

The difference between this and the outcome measure, this is seeking to measure the entire population that has depression or dysthymia, regardless of what that PHQ-9 score is. Simply a way to ensure and promote frequent use of the PHQ-9 tool.

This measure has been collected in the state of Minnesota as part of a suite of measures. It's also included in CMS' Meaningful Use Program. Our average statewide rates on over 100,000 patients is 68 percent. I'm sorry. I'm a little bit out of -- I was

planning on presenting the other measures first.

The statewide average is 65.6 percent, with significant variability among the clinics. Some of the data that was presented to you, one of the issues that the work group had was it seemed like this was not moving very much. We reported rates at a medical group level.

But even within a large, integrated medical group, for example, one group has a range of clinics administering the tools between 43 percent and 98 percent within the same integrated system. So there is opportunity and room for improvement.

CO-CHAIR PINCUS: So Dave, do you want to sort of begin to go through the importance to measure and report.

MEMBER PATING: Yes. Thank you very much. Well first of all, I'd like to just announce that I have no conflicts of interest. I have no opinions about this

measure, and so --

(Laughter.)

MEMBER PATING: I'll put the force shield around me, yes. So this measure is again looking at the PHQ-9 and what was been around, a measure that's been around for a long time, originally tested as part of the PRIME-MD, and a standardized measure of using many systems across the country for depression.

The numerator is that you've done one PHQ-9 in a four month period, if you have a diagnosis of depression or dysthymia, and the documents and the folks in Minnesota were saying that by national count, 6.6 percent have a major depression diagnosis in the last year, and then if you add dysthymia, we're about up to 9.1 percent of the population.

I'm just going to be very brief in terms of the Section 1A, B and C. You know, depression is common and measuring improvement is felt to be a gap, or that many people

because of either being untreated, inappropriately or inadequately treated, there is a gap in people that have been diagnosed getting to the goal line of being, I guess, undiagnosed, because they've been well.

So there's felt to be again a commonness, a gap in treatment and then moving people that have been diagnosed to remission, and that the overall cost of this, in terms of life and quality of life and other health measures is significant.

So that's really the reliability issues and excuse me, the evidence issues. There was a lot of agreement among our group as to the basic need and the gap.

CO-CHAIR PINCUS: Other comments on the importance to measure and report?

(No response.)

CO-CHAIR PINCUS: Okay, wow. Okay. Seeing none, why don't we move to voting?

MS. BAL: Okay. Voting for

evidence is now open for 0712.

[VOTING.]

CO-CHAIR PINCUS: And Mike, I think, is conflicted.

FEMALE PARTICIPANT: Yes.

CO-CHAIR PINCUS: Okay, no. I want to make sure that that was said.

MS. BAL: Actually, I just do need one more vote. We should be at 23.

(Pause.)

MS. BAL: Okay. So then the results for evidence is high 21, moderate 1, low 0, insufficient 0, insufficient with exception 0 for 0712. And do you guys want to discuss --

CO-CHAIR PINCUS: Gap. Vote.

MS. BAL: Okay. Voting is now open for gap for 0712.

[VOTING.]

MS. BAL: We're missing one. Could everybody make sure that they voted?

(Off mic comments.)

MS. BAL: Okay. So for gap for 0712, we have high 20, moderate 3, low 0, insufficient 0, and we can move forward to high priority.

CO-CHAIR PINCUS: And high priority.

MS. BAL: The voting is open.

[VOTING.]

MS. BAL: Okay. So for high priority for 0712, we have high 19, moderate 3, low 0, insufficient 0, and we can move forward to scientific acceptability.

CO-CHAIR PINCUS: Great. So you discussed the scientific acceptability, reliability and validity.

MEMBER PATING: So this was done in Minnesota, and I actually forgot whether it was this measure or a previous measure that was done in response to a lawsuit. Was it this one that they implemented? Anyway, it might have been the other measure that we're looking at.

They looked at 80,000 claims. So it's very high utilization throughout their system, and they found a reliability of .846. I'm actually going to ask the submitters to, if they can explain this .846, because I wasn't quite sure. Was that a kappa or was it a Cronbach's, you know, alpha in terms of the reliability? So I'd be liking just a little clarification on that.

The group as a whole found using PHQ-9 to be on face value, you know, a very reliable measure for measuring depression, and again the goal of this measure is to document severity of depression when the diagnosis is made. So those are my comments.

CO-CHAIR PINCUS: Can the measure developer just respond to the question please?

MS. PITZEN: Sure. This is Collette. I just wanted to share. I know there was a little bit of difficulty technically with the documents and the insertions. But the reliability testing for

the measure itself was at .987, and that was part of the template that was submitted under 282.

And then the reliability for the PHQ-9, the PRIME instrument itself, had a sensitivity of 88 percent, specificity of 88 percent, an ROC analysis of .95 and Cronbach's also of .89.

MEMBER PATING: So in terms of reliability, I think we'd consider those very, very high.

CO-CHAIR PINCUS: Any comments on reliability?

(No response.)

CO-CHAIR PINCUS: Okay.

MS. BAL: Voting for reliability is now open.

[VOTING.]

MS. BAL: We're just missing one vote. If everybody could make sure to vote please. There we go, thank you. So the result for reliability for 0712 is high 19,

moderate 4, low 0, insufficient 0, and we'll move forward to validity.

MEMBER PATING: In terms of validity, that the measure measures depression and that you can tell something about it. There was strong belief that this was a valid measure. There were some questions around what would be considered the exclusions for the measure from the denominator.

The measure, as designed, excludes bipolar disorder and personality disorder. This actually goes back to the implementation of the PRIME-MD literature, which is from the mid-1980's, I believe, and they were found to, because of the high correlation of other symptoms related to those two diagnoses, the PHQ-9 was felt not to be valid for those two diagnoses.

It was mine and another reviewer's that there actually may be other diagnoses that we might want to recommend or consider in terms of exclusions, but I think the developer

will have a response to this. There's a sense that this may not actually be a valid tool, PHQ-9 in measuring depression in alcohol use disorders or substance disorders during an active phase.

I also wonder whether persons with cognitive disorders should be excluded, if there's a mood component as a part of a dementing or delirium syndrome of some sort. But there is a way that these are -- I'm sure that Mr. and Mrs. Developer, you could tell us how this works, but that you average out --

CO-CHAIR PINCUS: As you look up into the sky.

MEMBER PATING: --as you average out these various things. So could you address the exclusions and perhaps why these other exclusions, you felt, are not as significant?

MS. PITZEN: Sure, thank you. Appreciate the discussion today and at the work group call. We did conduct a literature

review for the use of a PHQ-9 for patients that have depression and alcohol or substance abuse, because that is not an exclusion from our measure.

The discussion that occurred at the work group, and we concur also that even considering the diagnosis of alcohol abuse or substance abuse might be difficult if one was to use that in a risk adjustment model, because of insufficient coding of that particular situation.

But we did thoroughly research the use of the PHQ-9 utilizing a bibliography from the tool developer. There were 1,340 studies cited as using the PHQ-9 or PRIME-MD. Searching on substance abuse and alcohol, there were 83 relevant studies that talked about using a PHQ-9 within this patient population.

Of those 83, only one study excluded patients with substance abuse. The rest of the studies included patients with

substance abuse, and there were several, actually seven supportive studies about comparing depression with and without alcohol abuse.

Probably the most significant one was by Delgadillo, Payne and Gilbody called "How Reliable is Depression Screening in Alcohol and Drug Users," and that was conducted in 2011. That demonstrated that with the PHQ-9 tool, that a PHQ-9 score of greater than 12 had a sensitivity of 81 percent and specificity of 75 percent for major depression and alcohol, displaying good retest reliability of .78 and internal consistency with a Cronbach's alpha of .84.

Within this population of people that were substance abusers, there was a 68 percent sensitivity, 70 percent specificity, and with modest retest reliability. So early on in the development stage, we had talked about the issues surrounding alcohol use and a patient's depression.

However, the work group at that time did not decide to exclude those patients from this measure.

MEMBER PATING: What about the persons with cognitive disorder, then, and then also -- can you explain the risk adjustment and how that may remove any systemic error then?

MS. PITZEN: I'll try. On the first question with the cognitive impairment, we do instruct our practices that if it is not appropriate to give a PHQ-9 to someone because of dementia or mental retardation, to not the use the tool.

One of the things that does is by not administering the PHQ-9 to those patients, you're not coming into the denominator for the outcome measures. In terms of the risk adjustment variables, I'm not sure I understand your question.

MEMBER PATING: Well my understanding is the way that you had

explained to us on the phone call that these variables like alcohol or cognitive impairment might be adjudicated somehow statistically in your risk adjustment of the measure, and I didn't quite understand that, and I was just wondering if you can explain the risk adjustment process.

MS. PITZEN: Sure, and maybe I lead you to some misunderstanding. So our current risk adjustment model includes the severity of a patient's depression. It includes insurance product as a proxy for socioeconomic status. It also includes age bounds of the patient.

So right now, we don't collect the existence of alcohol abuse with these patients, nor cognitive impairment. So we don't have those available for a risk adjustment methodology.

If they -- if some time in the future those are determined to be significant factors by the Development work group, we

would consider including that. In fact, we recently had a go-round about the impact of chronic fatigue syndrome, sleep disorder and some of those conditions, and those actually demonstrated that they were very, very low percentages within the population and they did not impact the outcomes.

So always a future consideration, but not right now part of our model.

(Off mic comment.)

CO-CHAIR PINCUS: Any other comments about validity? Rhonda. Rhonda, did you have --

MEMBER ROBINSON: Yes.

CO-CHAIR PINCUS: Okay.

MEMBER ROBINSON: I'm trying to get to it. I have a couple of questions about the PHQ-9. One has to do with the number of languages that it's been translated into and has it been tested for reliability in different languages?

MS. PITZEN: Hi, this is Collette.

So I believe the PHQ-9 is available in between 72 and 75 languages. I guess I am not the owner and developer of the tool, and I don't have the knowledge about the level of testing and reliability and validity. But those tools are available in different languages and have been translated as such.

MEMBER ROBINSON: Okay. Then my second question is really one around why one administration of the PHQ-9 in four months, and the selection of the four months, as opposed to a tighter time frame after the diagnosis?

MS. PITZEN: Sure, great question. Actually, we get the flip side of that question quite frequently, why isn't this measure like other measures? Why aren't you looking over a 12 month period for the administration of the PHQ-9?

This dates back to some of our earlier constructs for the actual ability to technically collect this measure. A four

month time period was chosen. So the purpose of the measure really is to support the outcome measures, which are looking longitudinally at a patient over time.

We encourage groups to be frequently administering the PHQ-9. You know, we can't dictate that you give that tool every time you see the patient. But the intent is that it be used to assess the patient's symptoms when they're being seen, and used as a means for follow-up contact with them.

So even this measure once in four months, in my mind you should be administering it more frequently as they're seeing that patient.

MEMBER ROBINSON: Yeah, I agree, and I think the reason why I asked that is that certainly this is an indicator of treatment and treatment effectiveness. The sooner that one understands the baseline and measures frequently, the more likely they're able to have a successful outcome.

So I think you've answered my question. I just think it's unfortunate that it's not requiring the PHQ-9 to be done earlier in the course, so that treatment could be more effectively administered, particularly if the person has side effects or is not responding in any way.

Data demonstrates within at least the first two weeks you can tell with the initial treatment whether or not they're starting to respond, and your earlier responders you treat a little bit more aggressively than you do those that don't respond, where you may have to change your treatment plan later on. So it's just a comment.

MS. PITZEN: This is Collette. Can I clarify a little bit?

MEMBER ROBINSON: Sure.

MS. PITZEN: So unlike the outcome measures, this measure is for all patients with that active or diagnosis in their

history. So one of the easiest ways to avoid being in the outcome measure is to never give the PHQ-9 tool at all. So we use this as a companion to evaluate that practices are implementing and using that PHQ-9 tool.

We have seen the rates of this particular measure soar. So the difference is this is the use of the tool. You could not be demonstrating a high PHQ-9. You could be in remission. But the point is to give the tool and do the assessment.

MEMBER ROBINSON: Thank you.

MEMBER PATING: Could I ask a kind of follow-up question and if I could step out of order, because it relates to the next two measures, at a six-month and a twelve-month.

What starts the clock as the index for that six and twelve months? Is it doing the PHQ-9, because you could be diagnosed at zero, and then do your PHQ-9 four months later, and then that makes your six-month measure really a ten-month measure?

So is it the diagnosis that will trigger the six and twelve, or will it be that when you do the PHQ-9?

MS. PITZEN: I'm sure -- this is Collette. It is when you're doing the PHQ-9. So actually we are looking for -- the diagnosis is a confirming factor, that we are dealing with major depression or dysthymia, because we wouldn't want to build a measure that was based on just the PHQ-9 alone.

But the act of that coming together, we call that an index. So the day that that occurs, the day that you have an elevated PHQ-9 and you also have the confirming diagnosis, that starts the clock ticking for every patient.

So every patient is going to have a different index date, and then that starts the clock ticking forward for six months plus or minus 30 day remission, and then assessing them again at 12 months, plus or minus 30 days for remission. So the same patient is

assessed at two different points in time.

CO-CHAIR PINCUS: So in a sense, this is a balancing measure, in terms of applying within the suite?

MS. PITZEN: That is correct.

CO-CHAIR PINCUS: Les.

MEMBER ZUN: The question I have is, and I'm not a psychiatrist obviously. Is the importance the delta of the PHQ-9 over time, or just getting a PHQ-9?

CO-CHAIR PINCUS: So Les, just to clarify. So you know, we're dealing in a suite of three measures, and this measure is -- and so we're talking about this measure and then we're going to talk about the remission measures at six and twelve months.

This measure is really trying to, as I understand it, it's trying to ascertain, of the universe of people who have a diagnosis that's applicable, how many of them have had a PHQ-9 within four years, to show that you're actually sort of reaching the population? So

it's not looking at a delta yet.

(Off mic comment.)

CO-CHAIR PINCUS: But the next measure will, okay. Other questions? Okay. So I guess we're ready to vote on validity.

MS. BAL: Voting is now open for validity.

[VOTING.]

MS. BAL: Okay. The results are high 19, moderate 3 for validity of 0712, and we can move forward to feasibility.

CO-CHAIR PINCUS: Okay. Dave, feasibility.

MEMBER PATING: So feasibility was easy. People felt that it was very feasible. While in some systems the PHQ-9 was not routine, people felt that the burden wasn't going to be huge, and then in the Minnesota data, they got 86,000 PHQ-9s. So that's 86,000 feasibilities that they did so --

(Laughter.)

(Off mic comments.)

MEMBER PATING: Yeah, somebody's doing good.

CO-CHAIR PINCUS: Any other questions that anybody has about feasibility?

(No response.)

CO-CHAIR PINCUS: Okay. I guess we're ready to vote.

MS. BAL: Okay. Feasibility is open for voting.

[VOTING.]

MS. BAL: Okay. The result for feasibility for 0712 is high 18, moderate 4, low 0, insufficient 0, and we'll move forward to usability and use.

MEMBER PATING: The group didn't have a lot of comments on usability, other than they thought this would be a good measure for quality improvement on an individual and a systems basis, as well as public report. Again, it's to measure who gets measured.

So using a measure was felt to improve quality, and then we'll see in the

subsequent when paired, that you'll get the delta and the change and the improvement over time. So we thought that there could be very good usability with a very validated instrument.

CO-CHAIR PINCUS: Any comments on usability?

(No response.)

CO-CHAIR PINCUS: Okay.

MS. BAL: Okay. Voting's now open.

[VOTING.]

MS. BAL: Okay. The results for usability and use of 0712 is high 20, moderate 2, low 0, insufficient 0, and we can move forward to overall vote. Unless there's further discussion, I'll open voting.

CO-CHAIR PINCUS: Any final comments or questions?

(No response.)

CO-CHAIR PINCUS: Okay.

MS. BAL: Voting's now open.

[VOTING.]

MS. BAL: Okay. The results are yes 22, no 0 for overall suitability for endorsement, and 0712 will be recommended for endorsement.

(Off mic comments.)

CO-CHAIR PINCUS: Terrific, okay. So now we're going to deal with these as a package. We're going to deal with the six month and 12 month ones, and let's hear from the measure developer, in terms of teeing this up.

MS. PITZEN: Great, thank you. As indicated, these outcome measures are identical, with the only difference is they're measuring the same patient at two points in time, assessing a patient for the patient-reported outcome of remission, the absence of depression symptoms as measured by the PHQ-9.

It's for adult patients aged 18 and older with major depression or dysthymia, and initial PHQ-9 score greater than nine, who

demonstrate remission at 12 months for the first measure, or six months for the second. And that is defined as a PH score of less than five.

It's a longitudinal measure that is looking over time to determine if the patient has remission from their depression symptoms. The beginning point for measurement is when the patient has both the diagnosis of major depression and dysthymia, defined by ICD-9 codes, and an elevated PHQ-9 score greater than nine.

This is considered the index, and we spoke about that earlier. Then we're looking longitudinally forward. Did the patient demonstrate remission at either the six month or the twelve month mark, and there is a 60-day window around that particular time frame, that we're looking for a repeat assessment with the PHQ-9.

The measure applies to both patients with newly diagnosed and existing

depression, whose current PHQ-9 score indicates that need for treatment. The measure additionally promotes ongoing contact between the patient and provider, as patients who do not have a follow-up PHQ-9 score are also included in the denominator, and they are assumed to be not in remission.

Exclusions for this measure include death during the measurement period, permanent nursing home residence and patients with a diagnosis of bipolar or personality disorders. In Minnesota, this measure is used in both primary care and behavioral health psychiatry settings.

The PHQ-9 tool is a patient-reported outcome tool with strong psychometric properties, and is validated for both aiding in the diagnosis of depression and for monitoring improvement of symptoms and assessing patient progress.

The measure is reporting on our consumer-facing website in a set of health

scores, and the 12 month measure has been selected by CMS for inclusion in Meaningful Use, and more recently for inclusion in the ACO GPRO Program. The measure that we just talked about was also included in Meaningful Use.

Some of the concerns expressed by the steering committee included low performance results, demonstrated by the measure, which has significant opportunities for improvement and the use of PHQ-9 tools for patients with the diagnosis of depression or alcohol or substance abuse, and we covered that issue already.

The measure is currently collected for all primary care and psychiatry clinics in Minnesota, the most recent data set representing over 350 clinics and 80,000 denominator patients. Granted, it's been difficult to see movement in the overall statewide average, which is currently at 5.6 percent, with higher-performing clinics at the

20 percent mark.

CO-CHAIR PINCUS: That's 5.6 percent for the six-month or twelve-month?

MS. PITZEN: At a twelve-month. The six-month's is a little bit higher. The six-month is 6.9. Not super-impressive, I know. But please note that for both the measures, the number of denominator cases have increased fourfold in the last four years, and a subsequent fourfold increase in the number of patients achieving remission.

It's well-recognized that maintaining ongoing contact with this population of patients with depression is critical to their successful remission of symptoms.

It's also very challenging to do so. Of any patient population, patients with depression are least likely to be able to self advocate and require processes and systems in place for maintaining contact.

Minnesota has made small,

incremental improvements in the rates of follow-up PHQ-9 at twelve months and six months, that are approaching 25 percent for twelve months and 30 percent for six months. The low outcome rates are not solely attributed to lack of follow-up.

Additional analysis of the denominator patients who do have a follow-up PHQ-9 at six and twelve months demonstrate that only about 25 percent of the patients are in remission, and another 25 percent demonstrate PHQ-9 scores between 15 and 27, which is severe depression.

We covered the alcohol/substance abuse issues and questions already. So thank you.

CO-CHAIR PINCUS: So just one other just clarification. It's my understanding that there's also -- there are other, two other parallel measures around significant improvement; is that correct?

MS. PITZEN: Oh yes.

CO-CHAIR PINCUS: But they're on a different time frame for reexamination?

MS. PITZEN: Right. They were recently endorsed in the last phase. We also have two intermediate outcome measures that are looking at the response for the patients, and that is a 50 percent improvement at the PHQ-9 score, again at six and twelve months.

CO-CHAIR PINCUS: So let's talk about importance to measure and report. Jeff, do you want to say something about the six-month and I'll say something about the twelve-month.

MEMBER SUSMAN: Yes. So I think we're already covered the idea that depression is an important issue, and that this is one that's worth concentrating on.

CO-CHAIR PINCUS: And even for a whole year it's worth concentrating on. So any comments or issues with regard to importance to measure and report, and the evidence supporting that, or with regard to

the gap, or with regard to the priority. Okay. So ready to vote. Now we're going to vote for the six-month.

MS. BAL: So we can only vote for one at a time.

CO-CHAIR PINCUS: Okay. So here's my proposal, is that we vote -- we essentially vote on the six-month at each of these stages, and then at the end, we go through the whole thing for the twelve-month, okay? Not the whole discussion thing, just the whole voting thing.

(Laughter.)

MALE PARTICIPANT: We'll be the judge.

MS. BAL: So this is the vote for 0711, which is the six-month measure, and since this is an outcome measure, we'll be voting on evidence slightly differently.

The options are yes or no for if the rationale supports the relationship of the health outcome to at least one health care

structure process, intervention or service. A slightly different decision, and voting is now open.

[VOTING.]

MS. BAL: Okay. The result is for evidence for 0711 is yes 22, no 0, and we can move forward to gap. The voting is open.

[VOTING.]

MS. BAL: Okay. The result for gap is high -- gap for 0711 is high 21, moderate 1, low 0, insufficient 0, and we'll move forward to high priority, and voting is now open.

[VOTING.]

MS. BAL: Okay. The results for high priority 0711 is high 21, moderate 1, low 0, insufficient 0, and we can move forward to scientific acceptability.

CO-CHAIR PINCUS: Okay. So let's talk about scientific acceptability. Jeff, do you want to make any comments there?

MEMBER SUSMAN: I think without

going into the detail, the Committee felt that these were generally reliable and valid. There were not major concerns. Of course, any measure can perhaps be improved. But this was as measures go, pretty darn good.

CO-CHAIR PINCUS: Yeah. I think a discussion around the twelve-month one, and I'm not sure whether this fits under reliability and measure specifications, or whether it fits under usability, is the question of whether it ought to be reported out as in different categories of people that were in remission, people who had significant improvement or whatever the term was, a response, and people who are not followed up, who did not --

So that that would provide a more refined way of reporting it. It seems like the data are there to do it, and it would be more informative to people. Any response from the measure developer on that issue?

MS. PITZEN: Sure, hi, this is

Collette again. We actually do have a suite of measures. We haven't put forth all of them for endorsement, but we are publicly reporting the follow-up PHQ-9s that accompanies the outcome measures. So of those patients who have an index and are included in the outcome measure, what is the rate of follow-up at six months and twelve months?

We report all of these measures as well. You can compare groups side to side. Like you could pull up the depression remission of the six months measure and the response measure side by side. But they are captured as separate measures, and we are processing --

The process is we're getting one file of information from our clinics and we're calculating all of the measures that you're talking about.

CO-CHAIR PINCUS: Okay. So I guess my comment, you know, doesn't go to the issue of reliability. It actually goes more

to the issue of usability, that it would be good to at least put that together, you know, in one place, so that individual provider groups could be looked at in that way.

But with regard to the specifications, you have the capability of doing that, you know, given the specifications that you have. Other comments on reliability? Rhonda.

MEMBER ROBINSON: Yeah. This is kind of a comment and really more of a question again. I'm just trying to play through the measurement of doing the PHQ-9 at four months, and let's just assume someone has started treatment as soon as they are diagnosed, and they don't do the first PHQ-9 until four months.

Let's assume that person has shown a response to the treatment and started on anti-depressant medications, and their PHQ-9 at that point may be nine or may be less than nine. So I guess what I'm saying is,

depending upon when the initial one is done, in relationship to the actual start of treatment, could very well skew these measurements towards patients who are, well let's see.

You're going to have a harder time if you use this measure and you delay the PHQ-9 measurement. You're going to have a harder time demonstrating those who are a large portion that would respond quickly and early if you do the PHQ-9 late, as your initial, and then when you do it again.

I'm not quite clear what you're getting. You may be skewing your populations. I don't know if I'm --

CO-CHAIR PINCUS: So I think Collette can respond to that, but I think it's also based on being above a threshold of the PHQ-9, to get into the denominator, and also -- this also incentivizes providers to do it early. But Collette, do you want to respond to that?

MS. PITZEN: Yeah. I can -- I'll try to clarify a little bit. I mean I've seen this a lot in the clinics and the records through validation.

Say you have a patient that maybe you're not quite sure if you have major depression yet, and you're administering the PHQ-9, and you then do that a couple of times before you give the diagnosis.

So that is the one confirming thing, is you need to have that diagnosis before we would accept a high PHQ-9 to start the clock ticking. But in terms of when you're doing the PHQ-9, I think we're probably mixing up the process measure and the outcome measures.

We encourage our groups to frequently, especially for patients that are identified as having an elevated PHQ-9, that you're maintaining contact with them and administering the tool on a frequent basis.

In terms of are they in remission

at four months, we did need to kind of draw a line in the sand, and actually we're looking for remission between five and seven months, is that time window around the six-month measure. So there's like a 60-day grace period.

But for everybody, the clock starts ticking at the same time when you have the diagnosis and you have an elevated PHQ-9. Does that help at all?

MEMBER ROBINSON: So the index has to be greater than?

MS. PITZEN: That is correct, greater than nine.

MEMBER ROBINSON: So the initial measurement has to be greater than nine, even if it's --

FEMALE PARTICIPANT: And a diagnosis.

MEMBER ROBINSON: And a diagnosis, which could be still at four months after treatment has started, could theoretically,

right?

CO-CHAIR PINCUS: Right.

MEMBER ROBINSON: So aren't you skewing this towards patients who are, how would you say, perhaps the more severe, and who are less likely to respond early?

CO-CHAIR PINCUS: It would be to a provider organization's disadvantage to do that. I don't know if Minnesota Community Measurement has any data reflecting on that. But I don't see why someone would purposefully do this, or systematically introduce this as an important factor.

MEMBER ROBINSON: I've raised my issue. It's just that patients respond at different -- those that respond within two weeks are different than those that are not responding until eight weeks and those that don't respond at all.

So I guess what I'm trying to figure out, if one is taking their baseline at four weeks, what is the patient population

that you're actually measuring at that time, of those who are less than nine?

They may have been -- if they did it before the treatment started, they may have been a 9, a 10 or 11. But you started the treatment. You didn't do the PHQ-9, and then they didn't respond. So when you measure it at four months, they already are not going to be in your cohort. So you're really getting those, if you started treatment right after diagnosis, but still delay the PHQ-9, then you're getting the more severe population.

MS. PITZEN: This is Collette. I'm going to try one more time. So, when we're talking about a time frame of four months, we're assessing anyone that has the diagnosis of depression, to just kind of see where they're at and make sure that the tool is being given to them.

But the difference between the outcome measure is it's the elevation of the PHQ-9 that is starting that process, and I

guess I can understand what you're talking about. If a patient is receiving active treatment, perhaps before a PHQ-9 is given. But it's been my experience that the PHQ-9 is really kind of starting that process and alerting that there's problems or symptoms. The second thing I wanted to share is we do risk adjust by the severity of that patient's initial PHQ-9 score, that's a part of their index.

CO-CHAIR PINCUS: Other questions about the validity? Tami.

MEMBER MARK: At the risk of sounding biased, I love this measure. I think this is like -- really like maybe the only, you know, true population-based outcome measure we have for mental health and substance use disorder, and it's actually being used widely and reported in a transparent way.

So I think it's quite amazing, but I might be delusional. So maybe that

perception is incorrect. So my question is I just want to make sure I understand, you know, what the data are showing us. Now this measure has been used for a number of years, and I think you talked to this a little bit. I thought what you said was that it's showing increases in follow-up rates.

But if you break out the people who were followed up, I wasn't sure what you were saying there, that the readmission rates have been going up over time, or have been flat over time. So if you can maybe just walk us through again what the trends are overall, and then broken out by whether the follow-up rates are going up since you've been using this measure, and whether the remission rates are going up, because since being used widely, and it's been used for a long period of time, you probably should, you know, assess whether it's moved the needle somewhat.

CO-CHAIR PINCUS: Collette.

MS. PITZEN: Go ahead. This is

Collette. Did you want me to respond?

CO-CHAIR PINCUS: Yes please.

MS. PITZEN: Like I had shared before, the follow-up rates for both the six month and the 12 month measure have been incrementally increasing over the years. The six month measure started in the low 20's, and now it's at 30 percent. But the 12 month measure, again it's a little bit harder to capture that patient one year out. But that measure has shown increases in the ability for groups to capture that follow-up rate.

But in terms of the -- when I'm looking at the patients that did have a PHQ-9, it's a special analysis that we do. We're not necessarily trending that over time. I know I have a suggestion. Why don't you just build your measure just for people that have followed up, and measure that?

Care and practice would not change, and we would still have 80 percent of patients lost to follow-up. But when we --

when I have done that analysis in the past, of the people that have had follow-up that six and 12 months, 25 percent of them are in remission. Another 25 percent of them are still at a level of severe depression.

So there's opportunity for patients feeling better, even in the ones that we are following up. So the poor performance of the measure can't be solely attributed to the follow-up alone.

MEMBER MARK: But so you could do a trending of the people who were followed up, you just haven't yet, correct?

MS. PITZEN: I have not trended that, no.

MEMBER MARK: That would be interesting to see.

MS. PITZEN: This is Collette. It would be interesting, but -- however we wouldn't want to promote that as the new measure.

MEMBER MARK: I get that. I

understand that. I just think it would be nice to know if the measure was moving the needle on improvement in depression outcomes.

MS. PITZEN: I think one thing that helps is that if we look at the numerator cases over the years, the number of patients that actually have achieved remission has improved fourfold. So we know, but we are making a difference.

CO-CHAIR PINCUS: So I think Tami's also touching on a broader issue that I think is important. We may come back to it again when we talk about the process of this meeting and our new role as a standing committee, is that sort of getting information about the performance of measures going to be really important as this committee continues.

So that we, number one, can have a better understanding of the context in which we're evaluating measures, and number two, can think of ideas for how to improve the measures, and number three, identify where the

gaps are, in terms of the measurement process.

So I think that that's something that, you know, we should think about in a more formalized way, in terms of how to get the information or feedback about the performance of existing measures, both in terms of their uptake as well as, you know, whether the needle is moving.

MEMBER SUSMAN: And I think to add just another brief thought about that, it seems to me if we can give feedback to the field in some of these measures, which are very strict, and I think, you know, it's really laudable in my mind that they have taken a strict approach to this.

But there may be a role for NQF, in demonstrating to the field that while we still aren't making the improvement in remission that we might, there are indications early on that we're at least getting people in follow-up a little better. We're actually doing PHQs, and you can't get to documented

remission until you do those things. So we're not there, but there is some at least early signal that we might be improving.

CO-CHAIR PINCUS: Okay. Dodi and David.

MEMBER KELLEHER: Mine is not a comment on the specific measure reliability or validity, but to follow up on a very simple thing that would make more sense to me is if we could in the future see the whole suite. I know off cycle it, but you know, it would probably have more meaning in terms of impact on quality if we knew, you know, saw it all the response and the remission at six months, response and remission at 12, and having it off cycle. It's sort of a little disconcerting for me.

CO-CHAIR PINCUS: Good point. David.

MEMBER EINZIG: So a question, just to make sure that we're comparing apples to apples in terms of clinics and providers.

Are we separating out the locations and other variables, for example, urban clinics, versus upper middle class suburban clinics, versus Indian Health Services, or rural areas? Because I would imagine there would be differences.

CO-CHAIR PINCUS: So there is risk adjustment built into this measure, and but Collette, do you want to respond to that?

MS. PITZEN: Sure. So for our public reporting website and our consumer-facing website, we typically will do some stratification by specialists, versus primary care providers. So that is one option. We also, for another project that we do, we look at things in terms of health care disparities, but we typically are not -- it would be interesting, but we currently are not doing that kind of stratification in our reporting.

CO-CHAIR PINCUS: Peter.

CO-CHAIR BRISS: And since risk adjustment in general is a hot topic at the

moment, can you review for us the variables that go into the risk adjustment model, and how much different they made?

MS. PITZEN: This is Collette again. I'm not a statistician. I'm trying to find the information that we provided in terms of that. I'm sorry. I'm just looking through my notes here.

MEMBER SUSMAN: There is an extensive description, I think, beginning on page eight of the document and going through the risk model.

MS. PITZEN: Right. So the variables that we're looking at, we're looking at age bands 18 to 25, 26 to 50, 51 to 65. Gender was evaluated. We're looking at the severity of the depression, based on the initial index, and then our --

MEMBER SUSMAN: And also payor.

MS. PITZEN: Pardon?

MEMBER SUSMAN: And also payor.

MS. PITZEN: Actually, that is

captured through insurance product, which is a roll-up of payors, and that is part of the risk adjustment model as well. So to just quote some of the statistics, the Pierson correlation compares the risk-adjusted and unadjusted depression outcome rates as .95, showing a strong correlation between the unadjusted and adjusted depression measures. Kendall's tau correlation was at .81, still strong but not as strong as the .95 with the other method.

So our statistician and the group that's working on this felt that this was a reliable risk adjustment model.

CO-CHAIR PINCUS: Other comments or questions about reliability or validity? Okay. So I guess we're ready to vote. So we can go through the reliability and validity voting components.

MS. BAL: Okay. Voting's now open for reliability of 0711.

Okay. So the results for 0711

for reliability is high 19, moderate 3, low 0, insufficient 0, and voting for validity is now open.

So the results for 0711 validity is high 18, moderate 4, low 0, insufficient 0, and we can discuss feasibility now.

CO-CHAIR PINCUS: So with regard to feasibility, Jeff, do you want to add anything?

MEMBER SUSMAN: I think the results speak for themselves. Tens of thousands of participants across the state.

CO-CHAIR PINCUS: I would agree. You know, it's not easy to do it, but it can be done and you know, it takes leadership to do that.

CO-CHAIR BRISS: And the other thing about that is that it's been done in tens of thousands of people. But the follow-up rates are still like at 30 percent, right? Isn't that what we heard?

CO-CHAIR PINCUS: So I guess the

point being that it's not so much that the measurement's a problem; it's the getting people to actually do the clinical care is the issue. Any other comments with regard to feasibility? Okay.

MS. BAL: Okay. Voting is now open for feasibility for 0711.

Okay. The results for 0711 feasibility is high 16, moderate 7, low 0, insufficient 0, and now we can speak about usability.

CO-CHAIR PINCUS: With regard to usability and use, I think we've already heard some comments about displaying it in a way that you could see the different categories sort of lined up. But any other comments beyond that?

MS. BAL: Okay. Voting's now open.

Okay. The result for 0711, usability and use, high 17, moderate 5, low 1, insufficient 0, and we can move on to the

overall vote, unless there's further discussion.

CO-CHAIR PINCUS: Okay. Any last comments?

MS. BAL: Okay. Voting's now open for overall suitability.

Okay. So for the result for 0711, overall suitability, we have yes 23, no 0, and we can proceed forward with 0710.

CO-CHAIR PINCUS: So now we're going to go through all of the voting for the 12 month one, and before we do that, is there any comments? Do people have any issues they want to bring up in distinguishing between the six month and the 12 month?

MEMBER SUSMAN: Well, first I have two comments. One is I just want to ask, will there be at the staff level any harmonization with this in the NCQA measures, or they're really just two different sectors of the world, kind of measuring the same things in other depressions?

CO-CHAIR PINCUS: So we're supposed to have a harmonization discussion at the end.

MEMBER SUSMAN: I know. But usually harmonization is within the data, the set that we have, right? Or is the harmonization --

(Off microphone comments.)

DR. BURSTIN: Yes. We have the harmonization discussion scheduled. We have them scheduled for later today. If we don't get to them, we're going to do them on our December call. But it's not just between -- or within the measure, but it's actually different measure developers. We want to make sure that if the measures aren't competing, if they're just related, that they're harmonized to the extent possible so there are comparable results. So we'll definitely have that conversation.

MEMBER PATING: And then with regards to 12 months, I think there is a

question, you know, of the low -- I guess it's the change, the delta from the six months to the 12 months. So whether that is significant, is it enough just to measure progress at six months or do you want to put out a second measure again at 12 months, kind of move the system more incrementally?

I know that 12 month outcome is important, but just the data didn't show that there was a lot of movement from measuring at six to measuring at 12. So I guess I don't know. I don't even know where it fits. I mean on a practical level, did it make a difference to do it at 12? So that would be the question that I would ask.

CO-CHAIR PINCUS: I'm not sure we know.

MEMBER SUSMAN: Yeah. I was going say I'm not sure the state of our evidence is there, to be able to even answer that. My sense is that it comes to what is going to be most effective in the health system over time,

to move practitioners and systems to do things differently, and where does it matter?

I mean for example, for some patients who have more severe depression, we know certainly that you might have to go through two or three drugs. There might be a referral and other processes that would make it unfair to say at six months has the person reached remission, or might be anticipated from the data we do have, that it would be a better measure of total outcome to say at 12 months.

CO-CHAIR PINCUS: I mean my view is that sort of the secret sauce in depression care is ruthless follow-up, and I think this encourages that. Any other comments, questions? Okay. So now we're going to go through, you know, the voting process for the 12 month one.

MS. BAL: Okay. So for 0710, evidence is now open, and the options are 1 yes, 2 no, since this is an outcome measure.

All right. So for evidence for 0710, we have yes 22, no 0. Okay, now voting for -- sorry, wrong one. Ignore that. Voting for gap is now open for 0710.

Okay, and the result for 0710 for gap is high 23, moderate 0, low 0, insufficient 0, and voting for high priority is now open.

Okay. The result for high priority for 0710 is high 22, moderate 1, low 0, insufficient 0, and we can vote on reliability now. We are waiting for two more votes, so everybody please make sure to vote.

Okay. So the vote for reliability for 0710 is high 21, moderate 1, low 0, insufficient 0, and voting for validity is now open.

Okay. So the result for 0710 validity is high 19, moderate 3, low 0, insufficient 0, and we can vote for feasibility now. If everybody could just make sure to vote, please? We're missing a few.

Okay. The result for feasibility for 0710 is high 16, moderate 6, low 0, insufficient 0, and we can move forward with the vote for usability and use is now open.

Okay, and the vote for usability and use for 0710 is high 19, moderate 4, low 0, insufficient 0, and now the vote for overall suitability is open. If everybody could please vote. We are missing a few.

Okay. The result for overall suitability for 0710 is yes 23, no 0, and this measure will move forward for endorsement.

CO-CHAIR PINCUS: Terrific. So now we're on our last measure to consider.

(Off microphone comments.)

CO-CHAIR PINCUS: Yeah, for this set. No, we don't have anymore. No, what I'm saying we're done. Now we're on the -- yes. Now we're on the last one. Yeah. You scared me for a minute.

(Off microphone comments.)

So this is -- so this is Measure

No. 2620, Multidimensional Mental Health Screening Assessment, and we have the measure developer here to tee it up.

DR. DAVIS: Thank you. Can you hear me okay? All right, great. Hi. So I'm Steve Davis with M3 Information. So our goal with this measure, and we've been through several iterations and, you know, there were several of you who were on the call, I think, a week or so ago, and discussed some concerns about how we wrote it up.

I'm a practicing psychiatrist. This is the first time I've written one of these things for you guys. I've done it on EURAC, but not over here. So you know, I appreciate being gentle with me, if you can.

(Laughter.)

CO-CHAIR BRISS: We haven't too much of a capacity for that.

DR. DAVIS: That's why I asked that question. So one of the challenges that we see in primary care, you know, as all of

you know, primary care providers do the lion's share of treating people with mental health problems, depression, anxiety disorder, some milder bipolar disorder, some mild to moderate PTSD, drug and alcohol abuse.

And those numbers range from 50 to 75 percent of people doing that, and they don't have psychiatrists in other -- or psychologists, other people to help them. They are not using tools, other than maybe the PHQ-9. In fact, just in the past month, I interviewed about five FQHCs about this, and most -- almost all of them were doing the PHQ-9.

None of them were checking for bipolar disorder, and we know that 20 percent of people who screen positive for depression have bipolar disorder. In fact, even in the study from the University of Pittsburgh last year that was in JAMA, about looking at screening 10,000 women, postpartum women for depression using the Edenberg, and then doing

a SCID on the ones who scored positive.

Twenty-two percent of them had bipolar disorder based on the SCID, very high. So by just using the PHQ-9 -- and I understand that these last Minnesota ones excluded people with bipolar disorder. I didn't see how they excluded them. I don't know what they did to make sure people didn't have bipolar disorder.

But the challenge is what everyone's doing is they're pretty much doing PHQ-9 and figure okay, we're done with behavioral health, and they're not. In fact, they're guaranteed to be mistreating some fifth of their patients who score positive for depression.

Anxiety disorder is twice as common as mood disorders, and are highly managed in behavioral -- in primary care practices, but there's no systematic way of assessing that. Nobody's doing it. How are you doing? Okay, fine. Here's your Xanax or whatever.

So we set out to have a multidimensional mental health assessment tool that assesses the common domains that are seen in primary care practices, and mood disorders including bipolar disorder, anxiety disorders including PTSD, drug and alcohol abuse.

Screening questions. Now this is a process measure. So let me say something about that, because in my initial application to this, we got it kind of, I think, conflated between process and outcome. I think we've cleaned that up pretty well, and it's really about the process measure.

And that is, I think, a bar at which to start, because this is not being done at all now in primary care practices, and patients are not getting screened and adequately assessed, a screening tool like ours or any others.

I mean there's the full PHQ. I haven't seen anything about the PHQ, the full 28 question PHQ for years, which looks at a

number of different domains. Nobody's using it, I think because it takes too long and it's not practical.

But there are practical tools. We've developed one that can be used in outpatient settings by primary care doctors, and we've added the proviso, and hopefully -- well, you have on screen, you have the most recent revision of this, which includes measuring this in practices that have staff-supported care, that can help in further diagnosis, referral and management.

Like what's added in the U.S. Preventive Services Task Force for Depression Screening. They say well, you should only do it if you can do something about the result, and we're saying the same thing.

So that's kind of, I guess, the basic thing that we're trying to do here. I'll just point out a couple other things that I think are -- I think I heard Peter, you talked about let's vote for absolutely

shocking.

Well, some of the absolutely shocking things to vote for are that a third of patients, only a third of patients with these problems are receiving any treatment. Of the ones who are, a third of those are receiving minimally adequate treatment.

So we're talking about 12 percent of these patients getting minimally adequate treatment, which is just not acceptable and we could continue to do kind of single dimensional things, or we could do something broader, and that's kind of what I'm hoping to do.

The final point, I guess, is World Health Organization, top 15 causes of global disease burden in developed countries. Five of the top 15 are behavioral health, depression, bipolar, panic disorder, alcohol, schizophrenia. So we have to pay attention to this.

I think that in my work with

patient-centered medical homes, and in fact we've got NCQA has looked at us and we have the only NCQA recognition for a mental health measure in patient-centered medical homes.

They have recently added, as I'm sure many of you know, a greater focus on behavioral health, and when I talked to patient-centered medical homes, what they're focusing on is making sure that they get, you know, I have to do a behavioral health quality measure, so let's pick one. They picked depression. Unless we have something that's broader, they're just going to keep picking depression over and over again, and we're not going to get off the dime.

CO-CHAIR PINCUS: Okay. So Mike, do you want to --

MEMBER LARDIERI: Great, thanks, yes. So --

CO-CHAIR PINCUS: And let's focus on the, you know, the importance to measure and report this focus.

MEMBER LARDIERI: Right. So the work group identified that it was important to screen across multiple dimensions, for some of the reasons that the developer had identified. So we were pretty clear that yeah, you should -- it's important to do this, and then there's, you know, some issues around the gap.

There's a wide gap around screening in primary care for multiple behavioral health disorders, and we were pretty clear on that as well. There was some discussion that the -- well, that gets into the other thing. Some discussion about whether, as initially presented, it was to use the M3 specifically.

But I think this revision here is not specifically to use the M3, but to use the multidimensional screening tool. So that's been revised since we had our discussion. So gaps yes, and importance, yes, was the work group discussion.

CO-CHAIR PINCUS: So you want to

respond to -- anybody have any comments or questions with regard to the importance to measure and report?

MEMBER ZUN: I'm coming from a different angle on this. I think that doing this kind of analysis or survey of patients in an ambulatory practice is important. The one concern I have is we never really look at this overall, you know. There's so many different requests from so many different groups in an ambulatory setting that, you know, how do they have the resources and time and money to do all these things and make referrals?

So and this may be more an aside, but you know, I really think we should start looking at alternatives, like could we have a web-based program, and require them to do a web-based program, and to document that they did a web-based, rather than having people administer all these different tools.

So maybe that gets more into measures. But I think we need to start

pushing that agenda, rather than adding more and more surveys for primary care practices, ambulatory practices. Thank you.

CO-CHAIR PINCUS: Raquel.

MEMBER MAZON JEFFERS: I just had a question. Maybe it's part of the harmonization discussion, but is this being proposed instead of other separate screenings, or -- so we just voted on a host of measures using the PHQ-9 to screen for major depression. This multidimensional tool includes major depression. We also just voted on a host of other screening tools for substance use.

So I'm just trying to be clear. Is this -- the use of the multidimensional tool, is that being proposed instead of the separate tools?

DR. DAVIS: So I could comment on that. So the last three measures that you looked at only applied to people with major depression or dysthymia, not bipolar, not

anxiety, not other things. So I think it's a separate thing, at least compared to those three measures.

CO-CHAIR PINCUS: I don't know if you want to comment, the developer, about the web-based comment, because my understanding is this is web-based, as well as being able to give it to --

DR. DAVIS: Well, it certainly can be. Our screening tool, which we make available for free, is in fact web-based. It's also app-based, so you could do it with any browser. You could do it using an app. But that's about screening.

The comment, I think, was about treatment and the overwhelming requests on primary care providers, as far as making referrals and so forth. There certainly, this goes kind of beyond the measure. But there are certainly already a number of web-based behavioral health tools that are being used to manage patients, and to do kind of some lower

level treatment.

CBT online, for example. There are a number of those things that are either online now or coming online, and I won't get into the evidence about those. But those things do exist.

CO-CHAIR PINCUS: Jeff and Bob and -- oh. Vanita, you had yours up first? Okay. Well, Jeff and Bob and Vanita.

MEMBER SUSMAN: Could you describe, measure developer, just a bit more about what you believe the causal pathway is here? So I'm trying to link this with outcomes in patients.

DR. DAVIS: Yes.

MEMBER SUSMAN: So just strictly looking at the evidence basis as opposed to any of the other potential issues that might be raised.

DR. DAVIS: Right. So we know that, for example, anxiety disorders are highly prevalent in primary care practices.

Yet there's not screening for it, and if you're only using something like the PHQ-9, you know, if all you have is a hammer, everything looks like a nail.

So even people with anxiety disorders have mood disorder symptoms. But you're kind of blinded by not asking more specific questions. By doing that, and of course it's also the case that there's a lot of comorbidity.

So just because somebody has depression doesn't mean they do or don't have anxiety. There's a lot of overlap. So you can't just find one condition and say go home, we're done. By having a better sense of what their symptoms are, then you can tailor the treatment more specifically to that particular patient.

You can make better referral decisions, because if you find those 22 percent of patients with depression actually have bipolar disorder, and your practice

doesn't feel comfortable or doesn't have the supports needed to manage that, you're going to make a referral decision for that patient.

So those types of better decision-making would lead to better management, better outcomes. I think that without starting somewhere, and I think this is kind of a lower bar, starting somewhere to evaluate well, who's actually doing this, and I think Leslie what you said about throwing more surveys at primary care practices, that's absolutely right.

So right now they've got Medicare asked them to do depression and alcohol, because if you combine that into one, now you've got one instead of two. If you can do several in one, then that's less for them to do, and they can get more of it, the patients and the providers.

CO-CHAIR PINCUS: Bob.

MEMBER ATKINS: We thought about exactly the same issue, Medicaid health plans,

okay, around the country. We thought about exactly the same issue, what we say is good, especially from Medicaid numbers, dramatic over-reliance on the PHQ-9. When we have people with all kinds of serious mental illness, serious and persistent mental illness.

So we use the K-6 as a high level screener, for exactly the same rationale as you have, with the idea that if someone screens positive, we then do a substance use screen to catch the co-occurring, and then send them to a clinician and treat them, and try to make the referral. So we don't need to drop into the level of detail, because we're not the treaters.

I would think that as we have PCMHs and ACOs that are starting to accept population risk, that would be an exceptional use for this, because they'd have to find people and provide the treatment for the K-6. It's not -- it's too blunt an instrument for

that, and yours -- there may be others, but yours seems pretty -- a better instrument for that.

So I think that it really depends on the application, where it would return value to the entity that's doing the screening. I think that having it available for all the reasons you've talked about would be of tremendous value globally. But I don't know that it should be blindly said, well now we have to do this on top of that, on top of that.

I understand that concern. I think it's a realistic concern. But I think having it available as one of the critical tools in our toolbox is where I would see the value that this adds.

DR. DAVIS: Yeah, and let me just comment. You're right. ACOs and PCMHs are probably one of the types of groups that are much more likely to implement something like this, and the supports that are needed,

because they're on the hook for it.

So they're going to bring in behavioral health. They're going to make sure that they have programs that can track these, and you know, if you look at -- of course the excess cost in medical conditions, because of poorly managed or diagnosed mental health problems.

In Maryland Medicaid, for example, for 2011 adult data, we find that patients -- we looked at 10 different DRGs, and this is in your report as well, diabetes and CHF and so forth. People with comorbid, these are Medicaid patients -- comorbid mental health conditions had a hospitalization relative risk compared to those without, of two to four times across 10 different DRGs.

MEMBER ATKINS: Our data consistently replicates that.

DR. DAVIS: Substance abuse was four to seven times. If you had the trifecta, I call it the triple threat, a chronic medical

problem, mental health, substance abuse across the 10 DRGs, 18 to 15 times the hospitalization rate.

MEMBER ATKINS: Yeah.

DR. DAVIS: So we've got to do something about it, and thank you for your comment.

MEMBER ATKINS: Absolutely.

CO-CHAIR PINCUS: Vanita.

MEMBER PINDOLIA: I think my question is on the data. I'm just very confused, because originally this was specifically for M3, and that's where all the data -- And now I heard that they're -- and now it's not. So I think my question's going to come more when I look at that data portion.

MEMBER SUSMAN: Could I just clarify?

CO-CHAIR PINCUS: Larry is next.

MEMBER MILLER: I think this is a valuable tool, and I've got more of a technical question that sort of follows up on

Raquel's question about harmonization, and Les' questions about over-supply of process and surveys, screening tools.

And maybe this will come up with the harmonization, but I'm just sort of curious now for the staff or whatever. If we like this and endorse it, what do we do with some of the other tools that don't quite take care of all these global kinds of issues, and how do we deal with that then?

(Off microphone comments)

MEMBER MILLER: Right, I'm sorry. But if we endorse it, then we've got all these others.

DR. BURSTIN: Even prior to the endorsement decision, if there are competing measures, we'll bring that before you and you'll have an opportunity to discuss it. But for now, we ask you to look at the measures individually, on their own merit, and then we'll come to that.

MEMBER MILLER: All right.

CO-CHAIR PINCUS: So Peter and then Dodi.

CO-CHAIR BRISS: So I had a -- so I don't recall seeing any actual data presented about the relationship of using this screening instrument and the outcomes. I understand that you've made a conceptual case, that people are under-recognized. We think we have treatments for some of the conditions that you might recognize, and that ought to result in better outcomes.

So I think the logic model is clear to me, but I haven't seen any actual data sort of supporting that, the logic model. Am I right?

DR. DAVIS: So that's where I think things got confusing with the initial application, because we had a lot of data on our particular instrument. But the measure being submitted was a process measure, the percentage of patients in an adult population who have had this screening in the past 12

months, versus not.

And so we took our instrument data out of that, because it was really considered to be irrelevant to the process measure.

CO-CHAIR BRISS: But as a general -- as a general rule, if you're submitting a process measure, you know, you're supposed to make a case that the use of the process measure would result in better outcomes, right?

So at this point, I'm not asking you about reliability or validity of the measure itself. I'm asking about whether you have empirical data that links this sort of screening to better outcomes.

DR. DAVIS: Right.

CO-CHAIR BRISS: And especially in the context of -- I think we just rejected kind of a similar pediatric measure yesterday, because it didn't have a great link to outcomes. But it looked like -- based on the submission, I think it had more documentation

of outcomes than this one does.

So I'm just trying to -- I'm trying to get clear in my head, sort of, your relationship of screening and better treatments and outcomes, and I'm trying to make sure that we're handling, sort of conceptually, similar measures in a reasonably consistent way.

CO-CHAIR PINCUS: Steve, you can respond, and then Dodi and Jeff.

DR. DAVIS: All right, thank you. So in -- what we took out was sort of, for example, although I think maybe some of it, in the appendix there's still, I don't know, a five page document or so, that describes some data from the University of California at San Diego, where across four sites our particular measure, multidimensional mental health screening tool was used across four sites, 12 different physicians, and that data showed that all 12 providers found it to be useful.

I forget all the other aspects to

it that I had in there, but it should be in your appendix. So we've used it there. We have -- there's a large ACO in New York City that is currently using it. They are collecting data.

We did get some -- we did not get any specific numbers out of them, other than some anecdotal data. They had some pre/post measures with respect to suicide attempts in their population, and found that there were nine suicide attempts in a period of time. Prior to initiating this, zero; afterwards, again it's the anecdotal data.

CO-CHAIR PINCUS: So I just want to just clarify something. So you made a change in the measure, so that you're not specifying a particular patient-reported outcome measure, but that there be --

DR. DAVIS: Yes, and it calls for a validated multidimensional measure or tool.

CO-CHAIR PINCUS: And do you have a list of them that are --

DR. DAVIS: We do mention several. We mention ours. We mention the mini. We mentioned --

CO-CHAIR PINCUS: Right, right. You mentioned them, but what would be -- how would this -- how could this be uniformly applied across the country, so to speak, as -- in a reliable way?

DR. DAVIS: Right. So for example, we kind of looked at the model that's being used for depression now, which calls on users to use a validated depression screening tool. I'm wondering how it's different than that.

CO-CHAIR PINCUS: Well we just talked about using the PHQ-9.

MEMBER SUSMAN: Can I clarify just with staff what our task is, because I'm really confused about whether we're looking at the M3 or whether we're looking at the broader idea of any quote validated tool.

(Simultaneous speaking.)

MEMBER SUSMAN: I hear what you're saying, but I'm not sure --

DR. DAVIS: It's really just about the process measure. I don't know.

DR. BURSTIN: It has to be about the measure before you, not the broader issue. Whatever the measure is before you, what's included in it is what you need to make the decision on.

(Off mic comment.)

CO-CHAIR PINCUS: I assume that the measure has changed since the previous discussion.

MEMBER SUSMAN: Yes. I mean we had materials sent to us, and now I'm hearing -- so do you see what I'm --

DR. DAVIS: Yes. If I could comment on that. We've had -- when we had our conference call a week, two weeks ago, something like that, the committee who were on the conference call asked for specific changes to clarify our submission.

We worked for many hours with them, with Lauralei and Angela and Sarah, helping to make sure that we understood what was being requested of us.

I think that we answered that appropriately, and that's -- so we made the changes recommended by the Committee in what you have before you now.

MS. FRANKLIN: Steve, did you want to sort of summarize?

(Off mic comments.)

MEMBER SUSMAN: For Helen or Lauralei or whoever's the appropriate person, I mean, do we do something based on the developer that's just before us now, having not seen it, or do we use what was submitted and we all had the chance to review and I was on that small work group?

I guess we could go either way. my own sense is that having stuff come in the day of the review makes it very hard to give it appropriate due diligence.

MS. FRANKLIN: So Steve, if you could give us a quick summary of what exactly changed. I don't think any of the underlying material has changed. You have taken out the tool, the specific reference to the tool and kept -- retained the multidimensional aspect of the measure.

DR. DAVIS: Thank you. Yes, that is -- so we took out mention because, again, it was thought to be confusing about is this a process measure or is this an outcome measure. So we took out any of our --

(Off mic comments.)

DR. DAVIS: So we made it much more clearly around the process measure that we had defined. There was a comment as well about, well, just screening for something and not doing something with the results, not having staff to make referrals or better treatment, diagnosis and so forth was -- was not very useful.

So we added in the definition,

both in the numerator and denominator, language to make it clear that this applied to practices like ACOs and PCMHs, for example, that did have those staff supports to be able to manage the diagnoses that they come up with. Those are essentially -- I think those are the main points, changes.

CO-CHAIR BRISS: So now --

CO-CHAIR PINCUS: Let me sort of --

DR. DAVIS: I told you I'm a neophyte at this, and I probably screwed it all up, but I'm doing the best I can.

CO-CHAIR PINCUS: Well, let me see if I can -- so I think we're talking about two different issues here. One is we've been talking about the importance to measure and report this concept, this focus, and so there are issues there.

But there's a separate set of issues in terms of the lack of specification of the information before us, about what it is

we would be voting on, and I think that's a bigger issue. I mean we can talk about the concept, and I had my thing up because I had some concerns about the concept.

But I'm not sure we're in the position to be informed and to vote on what is the specification of this measure that we would be voting on.

MEMBER PATING: I feel like we just can't take eleventh-hour consideration. We really need to look at the documentation and consider these carefully. It's not a visibly fair process, but also I think there's still some structural concerns that really need to work out.

In the pre group, there was two issues. One issue had this concept paper, but you had no data backing the concept up, and you submitted specific indicator, the M3, which is the study that you've done.

We asked you pick one or the other. Either fill out the big picture or

give us the M3 indicator, and you actually I think took out the best part, which is the data around the M3.

But we still have this chasm, which is the concept, but no data, no specifications, no trial, no validity or reliability. So we're really stuck to go into this detail now, and I just feel like I can't even give this a vote.

I'd like to table this if we can, if there's such a thing.

CO-CHAIR BRISS: Yes. There's a thing called deferring actually, and so essentially we could defer to a later date, which I would recommend that we do. And so there are -- on the one hand you're to be commended, frankly, for trying to make a lot of changes in response to the work group call, right.

So but clearly, it made you and made us try to react to too many changes in not enough time. And so what I would

recommend is that we defer this one, which is essentially tabling it for a future time.

That would allow you a little bit of extra time to work with staff and possibly some members of the Committee about what would be needed, so that we could actually make a reasonably decision on this.

That saves us and you from -- it saves us from having to vote it down. It saves you from having to have it voted down and we could have a better discussion, I suspect, at a future date based on better information and more understanding of the process. So that's what I think I would recommend.

MS. DORIAN: We actually already -- no. We actually already have a post-comment call scheduled for January 8th, and oftentimes issues that weren't able to be resolved at the in-person meetings got discussed there, and measures are voted on again or for the first time if there wasn't

enough information. So that's a potential solution.

CO-CHAIR BRISS: I'd be careful about that, I think. This strikes me as being a possibly complicated measure. I don't see this one as being -- yes. My feeling is that if we were going to try to handle a measure by phone, I'd rather have it be a sort of a no-brainer, and my guess is that this one won't be a no-brainer.

But I think we ought to defer at this time, and talk about, you know, when and how we could bring it back up for a reconsideration is what -- I thought I saw a fair amount of head-nodding around the table. Are people generally okay with that? Is staff okay with that?

MEMBER PINDOLIA: One more comment. I think the discussion of having M3 be the specific tool versus any validated multidimensional tool, the data you had with M3 gave a lot of data. The concern was it

didn't address the substance abuse, which was a very important part of a multidimensional tool.

So I just don't want you to come back just with all M3 again, I think is at least my --

DR. DAVIS: Sure. Yes, we did have substance abuse data, but I guess we didn't present it or highlight it in a way that was helpful.

CO-CHAIR PINCUS: So assuming that we are now going to be deferring this, are there additional comments that people have?

CO-CHAIR BRISS: The other comment that I would make, as you think about how to bring this back is so in some ways, I consider the Preventive Services Task Force to be the gold standard in sort of making the case that some intervention or bundle of interventions improves outcomes, and that's a case that I didn't really see in the materials that you presented today.

So you might look to them for some examples of how they're marshaled evidence about what intervention am I talking about, and how do I know that if I do it, it will improve outcomes, and you might use some of those as examples.

DR. DAVIS: Thank you.

CO-CHAIR PINCUS: And so Dodi and David.

MEMBER KELLEHER: Yes. Just really quickly. I suggested on the prework call, and I suggest again that you might want to look at successful measures like the Minnesota Community process and then suite of outcome measures as sort of an example of what you need to get passed and endorsed.

DR. DAVIS: If I can respond to that with a question. I appreciate your feedback. So what I found challenging in doing that was comparing our process measure application with those outcome measure applications. I continue to get kind of

tripped up with the --

CO-CHAIR PINCUS: So number one is you have to decide --

DR. DAVIS: Is it just me?

CO-CHAIR PINCUS: Well, number one is you have to decide what the measure is that you're proposing, and then to think about what kind of evidence you need to marshal for that. You know we can't say whether, you know, it should be a process or an outcome measure.

But I will say, going along with what Peter said, that you know, as a process measure you'd have to document pretty strong proximal relationships with outcome, not a theoretical notion, but actually that there's a strong causal link between what's done in the process and there's a relationship with the outcomes, and not just, like I said, a theoretical notion, but actual data. So I think that's a key issue.

Okay. So I guess we now move into some final discussion topics. So do we want

to have the harmonization discussion now?

MS. DORIAN: I think maybe we'll do the other two first, I think. Angela? Or at least maybe we can take a break now and pick the terms, for a break between, because that's one thing we have to get done.

Additional Discussion Topics

MS. FRANKLIN: So our next activity would be picking terms for the standing committee going forward?

MS. DORIAN: Right. So with our new process of seating standing committees, as you've already heard --

(Simultaneous speaking.)

MS. DORIAN: That's true, right. Of having standing committees, it's good because for example on this instance, you're able to oversee the entire portfolio and review things on an ongoing basis.

So we have terms of two to three years that are randomly selected, and I had emailed everybody a while ago to say please

let me know if you have any opposition to being seated for a three-year term.

I didn't hear back from anybody, but if you do, you can let me know now, let us know now. If not, we'll just go around, and if you could just read -- say your name out loud and then read the results, so it's transcribed, recorded and transcribed.

(Off mic comments.)

DR. BURSTIN: All of which are renewable. So we'd love to --

MEMBER ZUN: And the answer is Les Zun, three.

MEMBER MILLER: I'm going to be in Gryffindor House. Three.

MEMBER EINZIG: Dave Einzig, two.

MEMBER PATING: David Pating, three.

MEMBER JENSEN: Lisa Jensen, three.

MEMBER ROBINSON BEALE: Rhonda Robinson Beale, three.

MEMBER KELLEHER: Dodi Kelleher, three.

MEMBER PINDOLIA: Vanita Pindolia, three.

(Off mic comments.)

MEMBER ZIMA: Bonnie Zima, three.

MEMBER MELNYK: Bernadette Melnyk, two.

MEMBER LARDIERI: Mike Lardieri, three.

(Off mic comments.)

MEMBER SUSMAN: You know, for a bunch of scientists, the fish bowl is very scary.

MEMBER GOLDSTEIN GRUMET: Julie Goldstein, two.

MEMBER MAZON JEFFERS: Raquel Mazon Jeffers, two.

MEMBER MARK: Tami Mark, two.

MEMBER KNUDSEN: Kraig Knudsen, two.

MEMBER SHEA: Lisa Shea -- oh.

Lisa Shea, two.

MEMBER SUSMAN: Jeff Susman, two.

MEMBER HORGAN: Connie Horgan, two.

MEMBER TRANGLE: Michael Trangle, another two.

MEMBER ATKINS: Bob Atkins, three. I'm a ringer. I belong over there.

(Off mic comments.)

DR. BURSTIN: Can we just explain to some people why we did twos and threes? It's just that -- so since we're starting the standing committee, we don't want all of you guys to rotate off at the same time. So the idea is just that half of you will be on for the two years and the other half, and it's still renewable.

So we're hoping that you'll actually come back for another two years. So after that, the terms will be two years. But at least we won't have everybody rotate off at the same time, and then we're stuck in the

same boat.

CO-CHAIR BRISS: Just to make everybody feel better -- sorry. Just to make everybody feel better, this side of the chair got the three. So we balanced it out a little.

(Off mic comments.)

MS. DORIAN: And the acceptance letters went out three months ago at this point.

So we have a discussion on gaps and sort of areas for future recommendations and more high level policy discussions, which we began yesterday.

CO-CHAIR BRISS: Could I make a comment? Yes. There were excellent comments on the gaps issue over the lunch break that I just wanted to put on the table. So we've approved most of the measures that we've seen over the last couple of days, but we approved nearly all of them with a bunch of important comments about ways that things, that might

evolve.

And so especially in a world of standing committees, it might be good to sort of try to figure out ways to capture what was said, and follow up with NQF and the measure developers about trying to more systematically evolve the measures, to kind of continuously improve the measures, so that, you know, over time maybe in a more systematic way than we perhaps have.

CO-CHAIR PINCUS: And it might be worthwhile actually, and maybe we should circulate this to the Committee. At the end of the last two Phase 1 and Phase 2, there were a series of gaps identified, with recommendations for future measure development, that we may want to sort of distribute to people.

(Laughter.)

CO-CHAIR BRISS: Do you always have that signal at the end of a meeting? You know, it's a good one for the group, right.

CO-CHAIR PINCUS: Dodi.

MEMBER KELLEHER: Yes. I want to second what you just talked about, and in fact maybe even recommend that there be a much more formal way that we document the sort of consensus recommendations or guidance, get that back to the developers, and maybe even have a strong expectation that unless those are addressed, and I think we --

CO-CHAIR PINCUS: It's almost like a study section, like you know --

MEMBER KELLEHER: Yes. If you don't address these, then your ability to get your maintenance endorsement may not be as strong, as an incentive for them to really pay attention to I think the very good comments.

In some of those gray zone sort of measures that got passed, I think that's especially important.

CO-CHAIR PINCUS: Vanita? Oh, Michael and Raquel and Vanita and David and Bob.

MEMBER TRANGLE: Can you hear me? Okay. I'd like to follow up, and Peter, this is sort of the comment or the question I was asking you during one of our breaks, but it wasn't in the general meeting.

It had to do with who are the parties that are really key or crucial to potentially look at the gaps, that just aren't part of our realm of thought or discussion? So you know, we've got our Joint Commission, you've got health plans.

But another major player that's especially important in the mental health and the substance use disorder world really is CMS, and what's going on, either that the feds are paying for with Medicare or MA, however it's configured, the role of the states, and some degree of accountability and sort of harmonizing that things kind of match.

CO-CHAIR PINCUS: Not to mention consumers.

MEMBER TRANGLE: And consumers,

you know. But -- because it just feels like we sort of talk about it, and then we kind of wring our hands and sort of go on to something we can deal with and take action with. The other thing I mentioned way back when, but I think it's come up here periodically is, you know, I think you guys --

Helen, you were very correctly and proudly kind of pointing out that we added 60 measures and reduced 60 measures, you know. But I think for us to sort of somehow in a more formal, reliably kind of actionable way is to look at what's the whole portfolio, and how does it play out in the real world of a primary care/behavioral health clinician?

When we're adding something, what's the burden of time and money that it might take, and factoring that into our discussions with feasibility would be good, versus just technical feasibility with an instrument, you know.

CO-CHAIR BRISS: At least just to

perhaps make you feel slightly better, you know, HHS clearly knows that this general area is a gap. There have been HHS people at most or all of this meeting sort of listening in, and HHS is the funder of this particular exercise.

So the fact that behavioral health is a huge issue and that sort of it still represents -- it still includes lots of measure gaps is not entirely lost on the HHS family.

CO-CHAIR PINCUS: No, but I think Mike is getting a little bit more specific than that, because I think what's important is that CMS is doing stuff, but we're not aware of it, you know. So that's -- you know, there are things that are in the pipeline from CMS or issues that have come up, for example, in the MAP, yes, in the MAP that --

Like, you know, we had a discussion last time about reviewing the measures for the health exchanges, and noting

the lack of behavioral health measures, and particularly about sort of access to specialty care.

CO-CHAIR BRISS: Would anybody else from the family like to comment?

MS. POTTER: Hi. I'm D.E.B. Potter. I've been here the whole time, except for when I had to run to another NQF meeting.

I work at AHRQ three days a week and at the Office of the Secretary two days a week, and I've been working with NCQA and Mathematica on these measures that Sarah presented to you, that were developed with money from the Office of the Secretary and SAMHSA, in partnership also with CMS.

So HHS is very aware of all that you're doing, and we take what you say and I write it down and I think about it. I appreciate all of your contribution and your volunteering to this effort. But knowing where the important gaps are is something that

helps the Department in general.

So you know, please make that known. I'll just leave it at that.

CO-CHAIR BRISS: And the other thing that I would say on this topic is that there's sort of in the -- among the people that are at HHS working on measure development or measure development and testing and the sort of other key players like NQF and NCQA and others, there's been an increasing amount of cross-talk.

We haven't reached nirvana yet, but there's an increasing amount of attempts to get the left hand knowing what the right hand is doing in a way that we perhaps none of us, public or private sector, have done ideally to date.

CO-CHAIR PINCUS: So Vanita.

MEMBER PINDOLIA: My comment is on I agree with what's been said about what to do for measure developers, to understand where our concerns were. But yet I think I would

challenge NQF to take maybe a different stance for those that are gray zone, maybe making the developer understand.

I know right now that's not what NQF does. But to take this back for NQF to consider, that they would be able to say these were approved in the gray zones, understanding those aren't their committees.

But if it passes, and then they have limitations of what they can be used -- and specifically understanding if they can be used for pay for performance, when we had so many concerns.

It passed between a 40 to 60 percent, and maybe at the next round, if they have everything resolved, it can then move to that next level. It just puts a real hardship on the providers and health plan, when we had all those concerns and they weren't, you know, discussed or -- they were discussed, but they weren't resolved.

CO-CHAIR PINCUS: Okay, Dave.

MEMBER PATING: I actually would like to second that, and actually I wanted just to say, in terms of this process, this is my third NQF meeting, and I just think actually this has been the best and most thorough discussions that we've had.

I actually want to commend our chairs for getting expert in leading us. We're an unruly group. But I also think that we're learning as a group.

This continuity idea, I do think that it's an advancement, because we're always finding we're bringing up other stuff that has been raised in other measures, and the continuity of this process over time has great value.

So I think it goes to what Vanita is saying, in terms of, you know, us remaining as a thinking body and taking a look at these measures. But I again want to commend our NQF leadership, our leaders and then everyone here for a wonderful meeting.

CO-CHAIR PINCUS: Thanks.

(Applause.)

CO-CHAIR PINCUS: Dave.

MEMBER EINZIG: So ditto. Thank you, everybody on the NQF, and if anybody has a Sharpie, if we can change this to a 3-0, I wouldn't be opposed.

CO-CHAIR BRISS: You can always re-up. You have plenty of time to re-up.

(Off mic comments.)

MEMBER EINZIG: And building off of what Michael was saying about representation at the table, especially as we're moving more towards integrative models of care, I'd love to see more social work/mental health case managers, maybe education representatives, as we are moving more towards putting psychology and psychiatry in the school systems. So let's make it a well-rounded conversation at this table.

CO-CHAIR PINCUS: Bob.

MEMBER ATKINS: A couple of sort

of disparate thoughts. In terms of some gaps, I see on here, I just scanned it quickly, there's something about recovery, and that certainly connects to me with recovery-oriented metrics, in terms of what consumers are looking for and also something I'm just recently sort of learning more about, is the difference between disease-oriented evidence and what people refer to as POEMs, patient-oriented evidence that matters.

So the stuff that matters in people's lives, rather than disease state metrics. I think that's a really interesting and important issue to look at, because so much we look at disease states and things that matter to us.

I also would ask us to think about the concept of community tenure, because we used this when I was at Magellan. I'm introducing it now in Aetna. But it's not something that seems to be in common use. It has to do with people that frequently readmit,

but they readmit longer than 30 days.

So they might have eight admissions a year, and it has to do with how long people stay out of the hospital on average between admissions. For many of our people with serious and persistent mental illness, that community tenure concept, but it's not well-defined. Every company defines it differently, it seems.

The other thing, and this goes along with another comment, I'd say, is there were a lot of sort of cases made that this is really, really important, and we have really bad care, and we want to put this metric out there because it's better than nothing, which I totally agree with.

My concern is with unintended consequences, that once these metrics go out there, there are people, for example, state government, Medicaid, leaders who are largely uninformed about the concerns that we talked about here, and they use them to create

incentives and penalties for health plans.

As all of us know who are clinicians, punishment produces behavior to avoid punishment. It doesn't produce the desired outcome.

So I'm really concerned -- I'm old enough to remember Skinner. I'm really concerned that for people doing -- we want these metrics for all the right reasons, but the unintended consequences of them are really scary to me.

And if we get follow-up, what -- how are they used, you know? Who's using them for what purpose? I think that's a really important consideration.

CO-CHAIR PINCUS: And Mike.

MEMBER LARDIERI: Thanks. I'm thinking that it would be helpful if we had, you know, for those consumers and other folks to participate, like the HIT community. So I would think it would be very helpful if we had some EHR vendors in, and if you look at EHR

vendors, you look at -- need to look at medical and behavioral health, because they're two separate groups, and they're not necessarily talking to each other now.

Or maybe from the Health Information Exchange area, and the other area under ONC, there's a Quality Measures Work Group, which is the work group that's talking about passing all the data. Once we decide what measure, they're figuring out how to pass that data around.

So somebody or some combination from those groups I think would be very helpful, because they're going to have to implement this stuff once we decide on it.

CO-CHAIR PINCUS: Jeff.

MEMBER SUSMAN: Helen, I think you hinted at this earlier. It seems to me there should be a higher standard for accountability measures, and I hear this in your remarks, Bob, about how either when we do pay for performance or we're looking at comparisons of

one plan versus another, I think many of the measures we talked about today seems like they will help to improve care.

Quality improvement efforts, I think, are really a very important part of our mission. I get real queasy about some of the measures we approve, because I think there is this transformation from improvement to accountability and potentially unintended consequences, where our evidence basis doesn't really, in my mind, connote.

And an easy way to operationalize that might be to use your gray zone sort of concept or something. That would be a two-stage process, where we vote for one level and then yet a higher standard level, or to rework your criteria around evidence and so forth, to be staged.

DR. BURSTIN: So I feel like Dave was just asking about the timing of that. We have a Consensus Task Force on the Board that meets actually next week, and this is one of

the proposals we're bringing to them. And after that, we will convene an expert panel to help us think that through.

It's interesting. There's not a lot of science that helps you figure out which criteria help you decide a measure for different intended uses. It's interesting, you know, when you speak to consumers and purchasers in particular, their perspective is, you know, in some ways if I'm going to use this measure to pick a doc for my family, that's just as important as how much the provider gets paid.

So you know, I think that's where you have to sort of think through those concepts. So part of what we've been thinking about is maybe not getting into the issue of saying this is for payment, this is for the different uses, but instead to almost have a -- I've been referring to it affectionately as sort of a bond rating.

You know, this is a Triple A

measure, this is a Double A measure. This one's still in its evolution, and sort of buyer beware, as opposed to necessarily saying use this for payment. But obviously when those decisions get made and you pick an A measure that's not been in use and isn't at the highest levels of reliability and validity, it should give pause, particularly when it gets to the MAP and they help select programs.

CO-CHAIR PINCUS: So you're talking about developing measures of measures?

DR. BURSTIN: You got it. Measure rating system, yes.

CO-CHAIR PINCUS: Okay.

MEMBER TRANGLE: If you think about that in terms of sort of the Triple AIM, you could almost sort of like do the measures by patient satisfaction, by quality and by cost/affordability, the three different perspectives, depending on who's looking at it.

CO-CHAIR PINCUS: Peter and then Larry.

CO-CHAIR BRISS: So this is a thought about process and proven. So I've been struck many times over the last couple of days about the divergence between what I think was in the -- what I thought I read in the materials, and what gets discussed around the table, right.

You know, so I won't pick particular examples, but it appears to me that unless my reading has gone completely awry in my old age, that there were some times when we had significant divergence. It might be worth thinking about whether there are process improvements that would allow this to be --

I'd feel better about the transparency and reliability of the process if you treated it more like a grant review, and if it's -- if the material isn't in print in front of everybody, it doesn't count, right?

CO-CHAIR PINCUS: Larry.

MEMBER MILLER: Well, this was my first rodeo in this arena, and when I got appointed, I was very happy, and then when I started looking at the data, I got very anxious. I didn't need a screen to tell me oh my God, what have I gotten into, you know. And so I know what I know, and there are other things I have no clue about, and I'm the first one to admit that.

So I listened very intently on the calls, and we were the first group to get assigned and I think we had about a week to do it. I go oh my God, and then I've learned a lot just sitting here and listening and sort of getting the process.

I think between Peter and Harold, I've really gotten a lot of input into sort of what this is all about for the next rodeo. I think in terms of some of the process, I think I like the idea of the work groups having more time to sort of process things and begin to talk about things, rather than just the one

call.

I know we're all very busy and I'm sure people won't be happy about it. But I think it helps the process to really think more intellectually about it, and really do a good job on those kind of things. So I would really be in favor of that.

The other sort of -- one of the other points is that I concur about the gray zone, because I think that even though they're documented, a measure is a measure when they get out there in many ways, and people don't know that all measures are created equal, and I think that's a problem.

One of the content areas that you may have addressed in previous iterations or not is the whole other spectrum of the population, the graying population, the elderly population in terms of behavioral health issues.

The population is aging, and I think that that would be another area that

many of us struggle with, in terms of the interface between what's intellectual, what's behavioral and how that all works together.

I think that would be an important piece to get, and I had a great time, thank you.

CO-CHAIR PINCUS: So Raquel, Tami.

CO-CHAIR BRISS: Mic, please.

MEMBER SIDDIQI: Oh, you can't hear me? Sorry. As we were chatting, I was chatting with Dr. Susman yesterday and Leslie. But my primary interest is behavioral health and elderly, and with all the regulations that are coming down with, you know, Department of Health and Medicare in terms of psychotropics in elderly, I have to say very honestly the tertiary psych facilities don't know what to do with these people.

Now they are stuck at our facilities like long term facilities or acute care hospitals. Really nobody knows what to do with these patients, and then there's

regulatory issues in terms -- and then there's quality of life issues.

I think there has to be a little better guidance, because they're regulations, but they don't make sense in certain populations. The intent is good, yes, but it doesn't apply to everyone. That's my dilemma. So I agree with you. Thank you.

CO-CHAIR PINCUS: Raquel.

MEMBER MAZON JEFFERS: I also want to say what an incredibly rich learning experience this has been for me. So thank you.

But I wonder if it's worth taking maybe a very small group of people, to look at the -- in terms of process improvement, to look at the documentations, the documents, because I found myself struggling and tripping over and re-reading and searching to find.

My guess is that there might be a way to streamline that. I mean eventually I found everything I needed. But it seems like

it might be -- it might be simpler. There might be a way to simplify it.

MS. BAL: And then just feedback, that we are in the process of updating those work sheets. Could you give a little more detail, like what exactly you found difficult to find just so we can know?

(Off mic comment.)

MS. BAL: Okay, thank you.

DR. BURSTIN: Poonam's our lean person, so she's the right person to work with on this. So we'll make sure that we get a group of you together.

CO-CHAIR PINCUS: Tami.

MEMBER MARK: Basically it was just said, yes. If we could --

CO-CHAIR PINCUS: There's a couple of things, as kind of homework. One is that we do want people to think about, and we're going to have this follow-up call in January. But people to think about it, and actually even before January, actually to respond to

the staff, I think, with regard to number one, recommendations for future measure development.

So this is -- what's up here is what was actually in the report from Phase 2, and obviously we didn't fill everything today in terms of those gaps, and to sort of look at that and think about how one might refine that, but also add to it based upon what you've observed and both here, but also what you observed back home as well.

CO-CHAIR BRISS: In addition to that, sort of -- so in every measure discussion I've ever been in, people come out with a very long list of additional needs.

It would be great if you have thoughts about the highest priorities. In addition to lengthening the whole universe, it would be good to give some thoughts about what are the highest priority things.

CO-CHAIR PINCUS: Right. I mean to really think about this, so that we just

don't add, but we actually, you know, do it in a refined way. Then also come back with a different set of recommendations, in terms of process improvements, what we had here.

A number of things have already come up, like the notion of having a different kind of template for patient-reported outcome measures that need to be transformed into performance measures, to think about how we frame the segmentation kinds of issues that we've talked about, in a way that, you know, can make the process proceed more efficiently for segmenting existing measures, you know, things like that, you know, to think about that.

And so you know, speaking for me and Peter can speak for himself, as he does, that it's really been delightful working with all of you and with the staff. Now we're supposed to hear from public comment.

CO-CHAIR BRISS: And while we're waiting for public comment, thanks. I'll add

my thanks to everybody, too. Thank you.

NQF Member and Public Comment

OPERATOR: Okay. At this time, if you would like to make a comment, please press star then the number 1.

(No response.)

OPERATOR: There are no public comments at this time.

Next Steps

MS. DORIAN: Okay, wonderful. Well, gosh. It's been -- these two days have actually gone by really quickly. We just -- we have one last slide for Next Steps. So of course this first one that we've just completed.

We will work on drafting the report with your recommendations, and being careful to concisely and clearly articulate the areas where we were in the gray zone so that we can solicit comments from members of the public and members of NQF.

That will be posted for a 30-day

period, from November 7th through December 8th. Then we do have a call scheduled, it should be in your calendars, to review and respond to those comments, and we may use that call, if there's any additional time, if we don't end up getting, you know, a vast amount of comments, to discuss maybe gap areas or something else. We can think about that as it gets closer.

We will then draft the report for NQF member vote. That's for a 15-day period through February 6th, and then we have our Consensus Standards Approval Committee review your recommendations at a pretty high level. They sort of review them as a whole group and pull out any that they wish to discuss.

Then we have endorsement by the Board, and finally an appeals process. You also did have an optional date held in your calendar for I believe this Tuesday, and that was if we were unable to get to all of the measures in time. So you can release that

date. We'll cancel it.

I'm sure you'll heartbroken about that. Don't cry too much. So yes. Just on behalf of myself and my colleagues, I'd like to say thank you so much. It's been a wonderful two days. I've really enjoyed getting to meet all of you, and I think it was a wonderful, rich discussion.

MS. FRANKLIN: I'd like to add my thanks to Lauralei's and thanks to our co-chairs for shepherding us through these two days, as well as all the hard work you all put in, and the good recommendations that you gave to everyone, including developers and NQF.

CO-CHAIR PINCUS: Okay. Thanks, everyone.

(Whereupon, the above-entitled matter went off the record at 3:12 p.m.)

E

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Behavioral Health Phase 3
Standing Committee Meeting

Before: NQF

Date: 10-02-14

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R Gross

Court Reporter