NATIONAL QUALITY FORUM

+ + + + +

BEHAVIORAL HEALTH STANDING COMMITTEE

+ + + + +

TUESDAY
FEBRUARY 28, 2017

+ + + + +

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

PRESENT:

PETER BRISS, MD, MPH, Co-Chair
HAROLD PINCUS, MD, Co-Chair
MADY CHALK, PhD, MSW, Treatment Research Institute
SHANE COLEMAN, MD, MPH, Behavioral Health Division Southcentral Foundation
DAVID EINZIG, MD, Children's Hospital and Clinics of Minnesota
CHARLIE GROSS, PhD, Anthem, Inc.
CONSTANCE HORGAN, ScD, Brandeis University
LISA JENSEN, DNP, APRN, Veterans Health Administration
DOLORES (DODI) KELLEHER, MS, DMH, D Kelleher Consulting
KRAIG KNUDSEN, PhD, Ohio Department of Mental Health and Addiction Services
MICHAEL LARDIERI, LCSW, Northwell Health
TAMI MARK, PhD, MBA, Truven Health Analytics
RAQUEL MAZON JEFFERS, MPH, MIA, The Nicholson Foundation
BERNADETTE MELNYK, PhD, RN, CPNP/PMHNP, FAANP, FNAP, FAAN, The Ohio State University*

BROOKE PARISH, MD, Blue Cross Blue Shield of New Mexico
DAVID PATING, MD, Kaiser Permanente
VANITA PINDOLIA, PharmD, Henry Ford Health System (HFHS)/Health Alliance Plan (HAP)
RHONDA ROBINSON BEALE, MD, Blue Cross of Idaho
LISA SHEA, MD, DFAPA, Butler Hospital; Care New England Health System
ANDREW SPERLING, JD, National Alliance on Mental Illness
JEFFERY SUSMAN, MD, Northeast Ohio Medical University
MICHAEL TRANGLE, MD, HealthPartners Medical GroupBONNIE ZIMA, MD, MPH, UCLA Semel Institute for Neuroscience and Human Behavior
LESLIE ZUN, MD, MBA, Sinai Health System

NQF STAFF:

SHANTANU AGRAWAL, MD, President and CEO
HELEN BURSTIN, MD, MPH, FACP, Chief Scientific Officer
JASON GOLDWATER, MA, MPA, Senior Director
ANN HAMMERSMITH, JD, General Counsel
KAREN JOHNSON, Senior Director
TRACY LUSTIG, DPM, MPH, Senior Director
MELISSA MARINELARENA, RN, MPA, Senior Director
ELISA MUNTHALI, MPH, Vice President, Quality Measurement
DESMIRRA QUINNONEZ, Project Analyst

KIRSTEN REED, Project Manager

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

ALSO PRESENT:

MARY BARTON, MD, MPP, National Committee for Quality Assurance
KYLE CAMPBELL, PharmD, MS, Health Services Advisory Group
DANIEL GREEN, MD, Centers for Medicare and Medicaid Services*
JAMIE JOUZA, MBA, CAPM, PCPI Foundation
JUNQING LIU, PhD, National Committee for Quality Assurance
TIFFANY McNAIR, MD, MPH, Centers for Medicare and Medicaid Services
GARY REZEK, West Virginia Medical Institute*
DAN ROMAN, National Committee for Quality Assurance
KATHERINE SAPRA, PhD, MPH, Centers for Medicare and Medicaid Services
ANDREW SAXON, MD, American Psychiatry Association*
EDUARDO SEGOVIA, PCPI Foundation
ANITA SOMPLASKY, RN, CHTS-CP, CHTS-PW, Quality Insights
SAM TIERNEY, MPH, PCPI Foundation
KERIANN WELLS, MPP, Mathematica Policy Research

JENNA WILLIAMS-BADER, MPH, National Committee for Quality Assurance

ALMUT WINTERSTEIN, PhD, University of Florida

*** present by teleconference**

CONTENTS

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

8:33 a.m.

DR. LUSTIG: Good morning. I'm Tracy Lustig. I'm the Senior Director for the Behavioral Health Committee. And I'm really glad to meet everyone in person finally and welcoming all of you here, our Committee Members, our Chairs, our developers who are here and will come, and everyone who's on the phone with us.

So, let's get the really important information out front. Restrooms, for anyone who needs them, right outside past the elevators and to your right. We have a really packed agenda today.

We are planning to have some breaks, we will hopefully stay on close to the schedule as we can. We will have lunch provided to the Committee Members only and staff only around 12:30 and then we'll have an afternoon break.

For anyone that doesn't know or hasn't figured it out, we do have the WiFi here. Anyone can log in using the user name guest and then the

password NQFguest, with NQF in capitals. We want to remind everyone on the phone and in person, if you could mute your line or mute your cell phones during the meeting. We do understand you may need to run out quickly for personal reasons, but please do that.

And one other thing that's not on here, we always like to remind folks to be a little bit closer to the microphone than you expect, so that everyone on the line can hear us and also get our transcripts. And so, I'd like to, before we proceed, turn it over to our President and CEO, Dr. Agrawal.

DR. AGRAWAL: Thank you. I won't make very long comments, I just wanted to welcome you all. This is actually my -- the start of the fifth week on the job for me. So, I still believe I have that new CEO smell, but we'll see.

And it's been a great really learning exercise to come to all of these meetings. I try to attend every one of them and stay as long as I possibly can. It's really wonderful to see all

of the work that you are engaged in.

This Committee is particularly large and I know we are really good at getting your volunteer time. NQF is really masterful at that, I've come to realize over the last five weeks. This topic, I have to say, is really near and dear to my heart, particularly the substance abuse side of it.

I got my start in health policy, working at a substance abuse policy center, particularly on illicit drugs. I spent a summer actually in D.C. at that time, working on tobacco legislation. This was summer of '98, when Congress was really taking on tobacco legislation, and it was a major reform.

And then, I also wanted to extend my thanks to Peter and Harold for their leadership. They've been doing this for quite some time, I mean, well before, I believe this is your fourth phase of work, but my understanding is they've been leading this work for far longer than that. So, thank you very much.

Peter is a great connection to the public sector for me. I'm just coming out of CMS, so it's really wonderful to hear -- I think wonderful to hear how things are going in the government and I'm very, I'm hopefully that things will ultimately work out well.

Harold, being at New York Presbyterian, I actually did my med school at Cornell and what people don't know, this sort of fancy Upper East Side medical school was right next to the freshman year dorm, there's a methadone clinic.

And so, you walk by it every single day, going to classes, and nobody realizes it until we did a third-year public health rotation, where I was assigned to the methadone clinic. And it was just eye-opening, it was great.

So, that's it. I think I just wanted to -- I know we have a few new Members, but I won't steal your thunder, so that you get a chance to highlight them.

I've looked over the measures under

consideration here, you have quite a long list, but, again, I'm really excited about them. I'm glad to see that we're addressing opioids, glad to see that we're addressing tobacco cessation. So, with that, I will turn it back.

DR. LUSTIG: Thank you. I like to hide behind being new too, except I'm here almost a year, so I can't do that anymore. Before we move to you, I'd like to introduce or have introduced the other members of my enormous team here at NQF. We have two other folks, Kirsten?

MS. REED: Good morning, everyone. I'm Kirsten Reed and I am the Project Manager for this wonderful project. And I'm excited to be here.

DR. LUSTIG: And hiding in the back, we have Desi. You want to say hi, Desi?

MS. QUINNONEZ: Good morning. I'm Desi Quinnonez and I'm the Project Analyst here on the project. I'm looking forward to a great meeting.

DR. LUSTIG: Great. And now, I'll finally turn to our Chairs to give their

welcomes.

CO-CHAIR PINCUS: So, hi, I'm Harold Pincus. And welcome back everybody that's been through this before and welcome to all the new Members as well. Delighted to be here. It's a long time in between meetings, and I know, if you're anything like me, I'm kind of like an Etch-A-Sketch at this point, where I have to sort of go through the process again to sort of remind myself about how to do this.

And I think it's probably the same for many of you who have been through this and it will also be somewhat new for the people, or will obviously be new for the people who are joining us for the first time.

There is a big agenda. There's some important maintenance measures to look at, in terms of whether they are still appropriate and meaningful in terms of not only the way the healthcare system has been shifting, but also in terms of the kind of evidence that we need to look at to renew those measures.

And then, there's a whole series of newer measures that we need to look at. And so, we have a lot of work ahead of us, but the Committee has been really a very superb committee in terms of the kind of substance of the discussions that we've had and the way in which the process has really moved through very well. So, I look forward to working with all of you.

CO-CHAIR BRISS: So, ditto. Welcome or welcome back. This is like our fourth cycle of this stuff, so it feels like a room full of old friends. I appreciate the chance to reconnect and I'm looking forward to the next couple of days.

DR. LUSTIG: Before I have Ann jump in, I realize that I was remiss, we do have some other NQF staff here I should let introduce themselves. Karen, do you want to say hello?

MS. JOHNSON: Good morning. I'm Karen Johnson. I'm one of the Senior Directors here at NQF. And because Tracy is still new, this is her first CDP, I was able to act as her buddy for

this project.

DR. LUSTIG: Great. So, I'll now turn it over to Ann.

MS. HAMMERSMITH: Thanks, Tracy. I'm Ann Hammersmith. I'm NQF's General Counsel. And for those of you who are veterans of our committees, you'll be familiar with our disclosure of interest process. I will go over it for new Members and also to remind our alumni of how this works.

So, when you were nominated to the Committee, you were required to fill out a rather lengthy form in which we asked you some very specific questions about your professional activities. What we do in the oral disclosures is we have you go around the table and tell us if you have anything that you would like to disclose.

Just because you disclose does not mean that you have a conflict of interest. Part of the reason we do this is for transparency purposes, so that everyone knows where everyone

else is coming from. So, you can tell us that you are involved in certain activities, but it does not mean that you have a conflict of interest.

Just remind you that you sit as individuals, you are not a representative of your employer or for anyone who nominated you for service on this committee.

One thing that's a little bit different about our disclosure is that we look for financial disclosures, but we also look for volunteer disclosures, only if they're relevant. So, you may have been on a committee for your professional society and you may have been a volunteer on a committee and we would look for you to disclose that, if it's relevant to the subject matter before the Committee.

It doesn't mean you have a conflict, but we would ask you to disclose that. I'm trying to see anything else. I think we're ready to go around the table. If you're on the phone, I will call on you at the end of the in-person

disclosures.

Just a reminder, please don't summarize your CV for us, only disclose things that are relevant to the subject matter that the Committee that will work on. So, I always start with the Co-Chairs. Please tell us who you are, who you're with, and if you have anything you would like to disclose.

CO-CHAIR PINCUS: So, I'm Harold Pincus. I'm Vice Chair and Professor at Columbia University Department of Psychiatry. Also, Director of Quality and Outcomes Research at New York Presbyterian Hospital. I'm also adjunct staff at the RAND Corporation.

I have been a consultant for both Mathematica and for Montefiore Medical Center. And I'm on advisory committees for a number of academic medical centers, but specifically relevant is both the American Psychiatric Association, I'm on their quality committee, and also on the Behavioral Health Advisory Committee for NCQA.

CO-CHAIR BRISS: So, I'm Peter Briss. And my day job's at CDC, although I think spend more time in this room, actually. I'm the Medical Director in the National Center for Chronic Disease Prevention and Health Promotion. I also direct the Office of Medicine and Science there.

I've been involved in consulting on a range of tobacco measures and, shockingly enough, we think tobacco is important. And so, I've done some consulting with CMS on 3229 and I've done a fair amount of popularizing of NQF 28 in the Million Hearts Initiative and other things.

MS. HAMMERSMITH: Dr. Beale?

MEMBER ROBINSON BEALE: Wasn't sure you were going this way. Okay. My name is Rhonda Robinson Beale. I'm a Senior VP and Chief Medical Officer for Blue Cross of Idaho. Disclosure, the only disclosure I have is the work I do every day, which is creating value-based payments and using performance measures in order to do that.

So, my perspective and my bias will be from the actual use of measures and the issues that come out when you're trying to tie that to a payment mechanism.

MEMBER CHALK: I'm Mady Chalk. I'm the Senior Policy Advisor to the Treatment Research Institute and a consultant on a number of contracts to CMS on the Innovation Accelerator Program and to ASPE on a variety of other things, many of which are related to measures. And I suspect I have a conflict, I think, with two of the measures that are being looked at today, so I will recuse myself on those measures.

MEMBER JENSEN: I'm Lisa Jensen. I'm an Advanced Practice Psychiatric Nurse, I work for Veterans Health Administration, just up the street at 810 Vermont, but I actually live in Salt Lake City, I'm a virtual employee. So, best of all worlds. I don't have any disclosures that I need to make.

MEMBER PINDOLIA: Good morning. I'm Vanita Pindolia. I'm the Vice President of

Ambulatory Clinical Pharmacy Programs at Henry Ford Health System's Health Alliance Plan. And I really don't have any disclosures.

MEMBER MAZON JEFFERS: Good morning. I'm Raquel Mazon Jeffers. I'm a Senior Healthcare Program Officer at the Nicholson Foundation in New Jersey. And I was formerly the Addictions Director for the State of New Jersey. And I don't have any disclosures.

MEMBER EINZIG: I'm David Einzig. I did the triple board training in Salt Lake City. I'm working at Children's Minnesota. I'm here to give mostly clinical perspective, I see a couple hundred patients a month, at least.

I work closely in the pediatric clinic, doing collaborative care models. I see mostly kids with co-morbid psychiatric and medical illnesses, so kind of a complex population. No disclosures.

MEMBER LARDIERI: Good morning, hi. I'm Mike Lardieri. And I'm from Northwell Health, I'm the AVP for Strategic Program

Development in the Behavioral Health Service Line. Been on some committees with Office of National Coordination for HIT implementation, EHR usability.

I'm on the PPS Committee under the DSRIP Program New York for New York City's PPS and Managed Care Organization Committee. And also, I'm on the clinical advisory committee for an EHR vendor, Core Solutions.

MEMBER ZIMA: And hi, I'm Bonnie Zima and I'm a Child Psychiatrist, Health Services Researcher, Professor in Residence, UCLA. I'm on the Committee on Research for the American Academy of Child and Adolescent Psychiatry. And I receive research funding from PCORI, the State of California Department of Healthcare Services, and Illinois Children's Healthcare Foundation.

MEMBER KELLEHER: I'm Dodi Kelleher. I'm an independent consultant and I work primarily with large, self-funded employers on designing integrated behavioral health and fill-in-the-blank programs, disability, medical, et

cetera. And I have no disclosures.

MEMBER SPERLING: Good morning. I'm Andrew Sperling with the National Alliance on Mental Illness. And I don't think I have any disclosures, other than, I'm an unfunded consumer representative to the National Association of Insurance Commissions. I'm not sure if that's relevant or not.

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

MEMBER GROSS: Good morning. I'm Charlie Gross, Vice President for Government Business Division Anthem, that's Medicare/Medicaid. No disclosures that I'm aware of, other than it's been many years since I took an advanced statistics course.

(Laughter.)

MEMBER COLEMAN: Shane Coleman, here from Southcentral Foundation, which is an Alaskan

Native owned and operated healthcare system in Alaska. Do some internal QA/QI and development of measures and things like that, but that's about it. Thanks.

MEMBER PARISH: Good morning. I'm Brooke Parish. And I'm one of the Medical Directors for HCSC, which is the Blue Cross for five states. My really only disclosure, and I don't think it is really a conflict, is I also work as a surveyor for Joint Commission and have been trained for a surveyor at NCQA. I'm also President of our district branch in New Mexico for the APA and an assembly member.

MEMBER TRANGLE: Michael Trangle. I'm a psychiatrist at an integrated health system in the Minneapolis-St. Paul area called Health Partners. My conflicts are, I -- they're all unpaid, I wouldn't say unfunded, somebody's paying for the initiatives, but unpaid, voluntary gigs for me.

But I chair every year a guideline for a regional quality place called ICSI, the

Institute for Clinical Systems Improvement, the Depression Guideline. And there are Minnesota Community Measurements for depression that I'm involved, so there was a revision of those measures this past summer and I led that work group. So, those depression measures, I need to recuse myself.

And I was also involved for the Quality Insights Measure that we're reviewing here and I'm recusing myself for the -- oh, I had it up here. It'll come up, it's the one sponsored by Quality Insights having to do with preventative screening for depression and follow-up.

MEMBER SHEA: Good morning. I'm Lisa Shea. I'm a psychiatrist at Butler Hospital in Providence, Rhode Island, where I'm Medical Director. I'm also a Clinical Associate Professor at the Brown University Alpert School of Medicine.

I sit on the Board of the National Association of Psychiatric Health Systems and on

its Quality Committee. And I also was a volunteer member of the HSAG TAP, where two of the measures are coming up that I will recuse myself from, 3205 and 3207.

MEMBER SUSMAN: I'm Jeff Susman. I'm a family physician, geriatrician, and Dean Emeritus at Northeast Ohio Medical University. I've the best job in the world right now, which is being on sabbatical.

(Laughter.)

MEMBER SUSMAN: And I don't have any disclosures. Thank you.

MEMBER HORGAN: Good morning. I'm Connie Horgan. I'm a health services researcher. I'm a Professor at the Heller School at Brandeis University and the Director of the Institute for Behavioral Health. My conflicts are that I am on the Behavioral Health Measurement Advisory Panel for NCQA.

My Institute also works as a subcontractor to Mathematica on the MIDS Project, funded by CMS. And my research is predominately

funded by NIH grants and I'm the Principal Investigator on the Brandeis/Harvard NIDA Research Center, that focuses on system improvement for substance use disorders. And we do a lot of work with performance measures and improving them, we hope. Thank you.

MS. HAMMERSMITH: Okay. Thank you. Is Bernadette Melnyk on the phone?

MEMBER MELNYK: Yes. Good morning. I have the flu very badly and I didn't want to expose you to that. I am Vice President for Health Promotion and Chief Wellness Officer for the Ohio State University. I'm also a psychiatric mental health and pediatric nurse practitioner.

I have one conflict to disclose. I have a company called COPE2Thrive that disseminates my evidence-based cognitive behavior therapy-based programs for children and teens.

MS. HAMMERSMITH: Okay, thank you. Thank you, everyone, for your disclosures. One more reminder, before I leave you, if during the

meeting you think you have a conflict of interest, if you think someone else has a conflict of interest, or if you think that someone is behaving in an extremely biased manner, please speak up in real-time.

We don't want to get two or three months down the road and then have someone say, you know, I think I had a conflict of interest. We'd like to know now.

You're always free to speak up in a meeting, in open session. You can also go to your Co-Chairs, who will consult with NQF staff. And you can always go directly to NQF staff and they'll discuss it with you and reach a resolution. Any questions or comments?

MEMBER ZUN: Yes. This is Dr. Zun. I'm on the way over from the airport and I did want to give my disclosure statement, if that's okay?

MS. HAMMERSMITH: Sure.

MEMBER ZUN: Well, in brief, I have nothing to disclose.

(Laughter.)

MEMBER ZUN: However, I am a Professor and Chair of Emergency Medicine with a secondary appointment in the Department of Psychiatry at Chicago Medical School. I am System Chair for Emergency Medicine in the Sinai Health System.

I also am President of the American Association for Emergency Psychiatry. At this point in time, we have not developed any standards, although we are -- one subcommittee is looking at that.

I do have a grant, actually a couple of grants. One from Teva looking at agitation and one from the Emergency Medical Foundation, again looking at agitation in the emergency setting. So, I think I've covered most of my involvements, but I don't see any conflicts with the measures being discussed today.

MS. HAMMERSMITH: Okay, thank you. Anybody else on the phone that we missed? Okay. Thank you, everyone. Have a good meeting.

DR. LUSTIG: Thank you. Before we move

on, I wanted to give my other colleague here a chance to introduce herself, who just joined us.

MS. MUNTHALI: Good morning. Thank you. And sorry for being late. My name is Elisa Munthali, I'm Vice President for Quality Measurement at NQF. And I just wanted to welcome you and thank you so much for being on the Committee.

DR. LUSTIG: So, this next slide here says portfolio overview. We're actually going to have a significant amount of time tomorrow to really take a step back and look at the behavioral health portfolio overall, where there are measures, try to identify where the gaps are and where we think we should be focusing effort.

So, for right now, we just really wanted to give you the overview of the measures that we'll be looking at today. Sorry, there we go. It's always technology. Since our call that we had a couple of weeks ago, we actually had some more measures added to this meeting.

So, overall, we do have 13 measures

we'll be looking at. One of the measures, we're not going to be voting on, we'll just be having more of a discussion on, and that's the ECHO measure. That will be tomorrow morning.

But of the others, we do have seven new measures. Two of them are eMeasure versions of some of the maintenance measures we'll be also looking at. And so, we're covering, as usual, I think a wide range of topics.

We have tobacco use and follow-up after ADHD medication, follow-up after hospitalization, some depression measures, alcohol use, opioid use, medication reconciliation, medication continuation measures. And so, we're really going to be covering a gamut of things.

We have a very busy day today, looking at ten of the 13, so we're going to try to get through that process very quickly. And I know we had a long call going over our criteria. It's always good to go over it once again, just to give you a refresher.

So, the NQF process, we do have major criteria and the specific processes for going through these criteria. First, again, we'll be looking at the importance to measure and report. This is a must-pass criterion and we have two different votes on this. One is on the evidence itself and then, one is on the gap.

Next, we'll be looking at the scientific acceptability of the measure. This is reliability and validity. Again, this is a must-pass criterion. So, we will discuss and vote separately, first reliability and then validity.

Next, we'll be looking at feasibility. Again, we want to cause as little burden as possible, consider alternatives. We will discuss and vote on feasibility, but feasibility is not a must-pass criterion.

And I guess I should have mentioned, when I say must-pass, if a criterion that is must-pass does not pass, we do not continue discussion, it ends right there and the measure is not -- and that does not pass. For usability

and use --

CO-CHAIR PINCUS: Tracy, one question.

DR. LUSTIG: Yes?

CO-CHAIR PINCUS: Even though we stop discussion, is the material that has been prepared, in terms of the reviewers getting feedback on it, does that go to the measure developer so they can potentially improvement or address some of this?

DR. LUSTIG: Yes. The measure worksheets, as they stand, actually have already been made available to all of the developers. So, after we discuss and vote usability and use, we actually have a final vote. And that's the overall, do we recommend this measure for endorsement?

And then, we will have some discussion, if needed, on related. We don't have any competing measures this time, which means we don't have to choose a best-in-class. We may need to defer some of that discussion to tomorrow, depending on how long our timing goes

today.

As a reminder, and you all have handouts for some of our criteria here, we have very specific algorithms that we use to decide whether it's rated as high, moderate, low, or insufficient.

For some of the measures, and we will be telling you this, based on our analysis, there are cases where high is not a possible rating. And so, we have -- so, the first one was for evidence. Next, we have an algorithm for reliability. And then, third, an algorithm for validity.

And one other thing to let you know is that, for our maintenance measures, we have had some changes to our processes here. These measures have been looked at already in the past. And for some of the criteria, specifically evidence, not necessarily gap, but the basis of the evidence and testing, there may not be anything new from the developer, and that's okay.

And the Committee, in those cases, may

decide to simply accept the prior evaluation and not need to revote on it. So, for example, if it's a maintenance measure, in the past they've provided systematic review of guidelines as the evidentiary basis, there's nothing new to that evidence base, there may be more and in the same direction, meaning it's even stronger evidence, we may choose to say, this passed before, we're going to pass again and not need to go through the actual process of revoting. But you don't have to not vote. We also --

MEMBER MARK: Can I ask you a question on that? The --

DR. LUSTIG: Yes.

MEMBER MARK: -- Maintenance, the standing work, so, there's no requirement that any work be done if something's submitted for maintenance? Meaning --

DR. LUSTIG: Oh, no --

MEMBER MARK: -- by work, I mean there's no requirement that they do a systematic review again or do -- I was a little confused

going through the material as to what exactly was required when someone resubmits something for maintenance.

DR. LUSTIG: So, we do expect that if there is -- we ask them if there is any new evidence since they submitted it. So, we do expect that they've looked to see if there's anything new. And so, those are the updates we expect.

But if the updates, if there was a guideline and the guideline has been updated, but the basis of it doesn't affect the evidence of what we're looking for, it only makes the case stronger, then we would say that we don't necessarily need to revote on it.

But that's also different than gap. The gap is something we do still vote on. But we can go through this as we're going through the actual process.

CO-CHAIR PINCUS: But I think, also, one of the reasons why we have experts around the table is -- if there's a claim that there's no

new information, maybe we'll identify the fact that there is new information.

MEMBER PINDOLIA: So, can I clarify? So, a measure that's coming up for renewal, there is an opportunity to share concerns if it should continue? Or no?

DR. LUSTIG: Yes, absolutely. Yes, we'll still begin with a discussion of every criterion and subcriterion. The question will be, at the end of that discussion, do people feel that there is a need to revote?

And in that same vein is what I think some of you are alluding to, in our processes, what we're saying is that, for a measure maintenance, we really should be putting more focus on current performance and opportunity for improvement, that's what we call gap, and also on usability and use, is the measure being used, has it had any impact, are there any unexpected findings in terms of the use of the measure?

So, what we're -- by saying, not needing to vote, is if people think that the

evidence, and the testing in particular, are solid and thought it last time and still think that, let's not spend time revoting and instead, focus more on the use of it.

CO-CHAIR PINCUS: You mean voting on that criterion?

DR. LUSTIG: Correct.

CO-CHAIR PINCUS: Yes.

DR. LUSTIG: Voting on that particular criterion.

CO-CHAIR BRISS: Yes. And so, as a practical matter, this is things like, if you believe that last time we discussed tobacco or depression that it was an important topic that there were treatments that those were underutilized, which I suspect we might be able to fairly quickly get to some agreement on. We don't have to spend a lot of time sort of relitigating that and we can move on to the other criteria.

DR. LUSTIG: Yes?

MEMBER TRANGLE: You mentioned earlier

about that we might talk about competing measures. And I also know that somewhere in our schedule later on is harmonization of measures. Is that going to mainly be talked about tomorrow and we shouldn't do it piece-by-piece here today? What's your thought?

DR. LUSTIG: Yes. So, we don't have a lot of related and competing measures. We -- and, actually, in this particular portfolio, don't have what we call competing measures, which are two measures that are head-to-head doing the same thing and we would need to pick the better.

We do have some that are related and that's where the harmonization discussion comes up. We have, I think, one measure that had a related measure that we'll definitely discuss today.

One of the things we had talked about was, we have a lot of smoking measures and there's question as to whether they are completely harmonized or if they're harmonized as much as possible. We will try to do that today,

I think, but it just depends on, we have a really packed day today, so we may defer that discussion until tomorrow.

So, I know we've got a lot of questions here, but are there any other questions before we move on to the nuts and bolts of voting?

CO-CHAIR PINCUS: One thing, you might want to remind people about the difference in the process for an eMeasure review.

DR. LUSTIG: So, actually, we have, I mentioned, two eMeasures today. And we actually have our resident expert in eMeasures, who's going to come up and explain all of that to us. And so, we'll probably go over it mostly then.

One of the -- the big thing, though, for the maintenance measure and then, now, the eMeasure version of it is, we don't need to revote on evidence, because the evidence piece itself is going to be exactly the same. So, whatever is voted on in terms of the maintenance measure, we'll automatically assign that vote to

the eMeasure, the pass or no-pass.

But I'm going to let Jason explain eMeasures, much better than I ever could. All right. So, with that, I'm going to turn to Kirsten, who's going to go over, like I said, the nuts and bolts of voting.

MS. REED: Yes. All right. So, the next couple of slides are just going to talk about how today is going to work and your role in today. So, I know this slide here, we've really gone over quite a bit in the last two webinars, so I won't discuss this again, unless anyone has any questions on this.

Again, we've gone over this as well, kind of your role as a Standing Committee Member. We've asked everyone to come prepared today, in reviewing the measures that you were assigned specifically, but also asked that you kind of familiarize yourself with all of the measures.

So, we will be having a brief introduction by the developers. You'll see a number of them are in the back of the room, so we

want to thank them for attending today's meeting. We will also be asking them to briefly introduce their measure as they come up for discussion.

There's obviously a place right here at the main table for them to sit down and they will kind of go through the measure and address any issues that were kind of brought up during the PAs and also your analysis.

Following their introduction, we will ask the lead discussants for that measure to kind of begin the discussion, provide a summary of any pre-meeting evaluation comments, and, like I said, emphasize any areas of concern.

The developers will then have the opportunity to respond to those questions. And then, we will open it up to the full Committee to discuss that measure and then, vote on each of the criteria before moving on to the next criteria. All right.

So, how does the voting work? Voting is by criterion in the order presented on the measure worksheet. As Tracy mentioned in her

discussion, the evidence, performance gap, reliability, and validity are all must-pass criteria.

And then, we'll go on to usability and use, feasibility, and the overall suitability for endorsement. If a measure does fail on one of the must-pass criteria, there is no further discussion or voting on that measure and we will move on to the next one. All right.

So, in front of you, each of you have this fun little clicker and this is what we'll be using to vote. We ask that when it becomes time to vote, you point your clicker towards Desi, who is over there by the window.

And when you do vote, the remote will kind of display what you chose in that little red area. You are able to change your response without duplicating it and only the last option pressed before voting closed will actually register in the system. All right.

So, for achieving consensus, again, I think Tracy covered this, but at least 60 percent

of the Committee Members who are eligible to submit votes must vote to make decisions. The Standing Committee has not reached consensus if the vote on importance to measure and report, scientific acceptability, or overall recommendation is between 40 and 60 percent.

If the vote on subcriteria under importance or scientific acceptability is between 40 and 60 percent, the Committee will vote on all criteria and a final recommendation. If the vote on either must-pass criteria is less than 40 percent, then there is no further evaluation.

And any measures were consensus is not reached will move forward to comment and the Committee will revote at a later time. Sorry, I have too many computers in front of me. All right. Any questions on how today will work? Yes?

MEMBER KELLEHER: You have six or seven discussants for each measure, who will be the lead-off if the developer doesn't sort of introduce it sufficiently? It's just different

from how it's been before.

MS. REED: Honestly, whoever wants to speak up, yes. Unless, do you want to assign a certain one for each measure?

DR. LUSTIG: What we planned is that someone could start off, one of the lead discussants, and then, what we would ask is that, other people can join in and voice concerns, but not necessarily repeat anything. We'd ask you just bring up new issues.

But, certainly, we're not even just limiting it to -- we expect the lead discussants will be the ones that really dug in the deepest, but if anyone else around the table has concerns about any measure, they should feel free to speak up.

MEMBER KELLEHER: Thank you.

CO-CHAIR PINCUS: And if nobody speaks up, we'll call on somebody.

(Laughter.)

MS. REED: Are there any other questions? All right. Then, I am going to pass

it over to Peter and Harold and they are going to begin the process for the first measure up, which is 3148. And do we have our developers with us for that measure?

DR. LUSTIG: While our developer is coming up, one technical thing I wanted to point out. People may be confused about new numbers, you see that we have measure 04 -- you remember 0418, what happens is that when an eMeasure version is introduced of a maintenance measure, they both get renumbered. So, measure 0418 is now measure 3148, but we left them both there to remind you.

CO-CHAIR BRISS: And let me try to set a couple of additional things in terms of managing the discussion. So, we'll try to do the -- if you want to take a turn on the floor, please let us know by turning your card up.

We came without our binoculars today, so it's a little hard to see everybody from a distance, so please do the best you can in turning your cards so that we can see them. The

-- try to be efficient in your comments.

As we've said already, we've got a lot of measures to get through today and we want to make sure everybody has a turn. And it's probably good if we try not to duplicate comments that have already been said. So, with that, I will get out of the way and we'll open with the first measure.

MEMBER LARDIERI: I have a question. Yes. And it has to do with the renumbering. There's a lot of effort across the country to keep these things straight and when you start to renumber, has anybody evaluated the cost of that? In EHRs, other systems?

I know, with my system, with some of the measures that we're using now, if you renumber them, I have to retool all over the place. So, I don't understand the purpose of just changing the number. Maybe we could talk about that.

MS. MUNTHALI: It's a great question. We -- when we renumbered, we were trying to make

a distinction between the paper-based and the eMeasure, the electronic-based measure. But we realize that the numbers are not really intuitive, so we are going through another effort to see how we can better align the numbers.

So, once you're looking at a claims-based measure that might have been in rural, you can see that this measure may be associated with it. But the consideration with EHRs is a good input as well.

DR. BURSTIN: And just to add to that, one question for you, as we've struggled with this, you want to make it clear they are different measures. They have different testing at times.

And so, one question, would it be better to have the same number with some indication, for example, that it's an eMeasure, as another potential way to look at this? We would welcome your input on that, whether now or later.

MEMBER SUSMAN: We were talking about

just kibitzing, why not just call it e3148?

MEMBER LARDIERI: Or just put an E behind it if it's an eMeasure? I mean, this 3225 that was 0028, I didn't notice that was an eMeasure. That's an eMeasure?

DR. LUSTIG: So, it -- and I think the conversation is a little confusing. So, the new eMeasure gets a new number, but the old measure gets renumbered. And I think that's where people are confused.

MEMBER LARDIERI: Yes, so maybe we could explain what 3225 is and in parenthesis 0028. It looked to me like it's exactly the same.

DR. LUSTIG: Yes, you're correct. It's an issue that we're having ongoing discussion here. But, yes, the old measure, 0028, because an eMeasure was introduced, the eMeasure has a new number and then, the old measure gets renumbered.

MEMBER LARDIERI: So, maybe we should just consider the cost that it's going to be for everyone to retool to rename your measure. And

then, also, if you look in an EHR and you're tracking that measure over time, how do I flip now in the EHR from 0028 to 3225 without a whole bunch of retooling? And that retooling is all across the country, not just one provider.

DR. BURSTIN: I'll just add, this is something that was an internal NQF decision, so it's -- if your advice is strong, I think it's something we can potentially look at fixing, putting it back to the old numbers and adding an E, which I always thought made a whole lot more sense. Harder from a numbering perspective, but we'll see what we can do.

CO-CHAIR BRISS: So, I'd like to suggest, this is a really important topic. It's clearly broader than just this Committee, so why don't we kick this back to staff to make some decision about and, if that's okay, we'll move on to the first measure.

DR. LUSTIG: So, thank you to our developers and please get started.

MS. SOMPLASKY: Thank you. Good

morning. We are pleased to introduce NQF 0418: Preventative Care and Screening: Screening for --

CO-CHAIR PINCUS: Could you please identify yourself?

MS. SOMPLASKY: Oh, I'm sorry. My name is Anita Somplasky. I'm from Quality Insights. We're a measure developer for CMS.

MS. WELLS: Hi. I'm KeriAnn Wells. I work at Mathematica Policy Research and led the testing effort for these measures.

MS. SOMPLASKY: Take two. We are pleased to introduce NQF 0418: Preventative Care and Screening: Screening for Depression and Follow-Up Plan, for consideration of NQF re-endorsement.

We will discuss two versions of the measure. The first is NQF 3148, the claims and registry-based measure, and the second is NQF 3132, the Electronic Clinical Quality Measure, or eCQM.

The claims and registry-based measure was first implemented in the Physician Quality

Reporting System, or PQRS, in 2008 and the eCQM was added to the Electronic Health Record Incentive Program, commonly referred to as meaningful use, in 2012, and is now included in the Quality Payment Program, or QPP, under MACRA.

This measure was initially developed and implemented in CMS Quality Reporting Programs to promote the evaluation and treatment of depression through routine screening. Substantial evidence indicates negative patient outcomes are associated with depression, making it critical to identify and treat depression in its early stages.

The intent of this process measure is that all eligible professionals screen patients 12 years and older using an age appropriate standardized depression screening tool and, if the screening is positive, a follow-up plan is documented on the date of the positive screen.

The measure requires that eligible professionals document a follow-up plan, for example, referral to a specialist such as a

psychiatrist or psychologist, to ensure that positive depression screenings signal a treatment plan. This measure focuses on children and adults and includes all visits during the 12 month reporting period.

This measure is supported by clinical practice guidelines, the United States Preventative Services Task Force recommendations for adults and adolescents, and the Institute for Clinical Systems Improvement healthcare guideline Adult Depression in Primary Care.

In young adulthood, major depressive disorder is associated with early pregnancy, decreased school performance, and impaired work, social, and family functioning. While primary care providers serve as the first line of defense in the detection of depression, studies show that PCPs fail to recognize up to half of depressed patients.

Based on the evidence and on reported performance rates, this measure reflects an important care process that clinicians are not

regularly providing. The average 2014 PQRS performance rate for this measure was 26 percent, with 7.5 percent of eligible professionals reporting.

Reported performance rates indicate substantial variation in both claims and registry data and demonstrate that many clinicians have the potential to improve the rates of depression screening and follow-up.

This measure has the potential to incentivize eligible professionals to detect more patients with depression earlier, which in turn improves community and population health and reduces cost.

In addition to establishing the importance of the measure and illustrating the performance gap across providers, we have also tested the claim and registry measure and the eCQM and found each to be reliable, valid, and feasible. Thank you for your consideration and we look forward to the Committee's questions and discussion.

CO-CHAIR BRISS: So, with that, would one of the lead discussants like to kick this off?

MEMBER SUSMAN: So, I think this is a measure where the evidence is largely the same, if not a bit stronger. The rationale has been well discussed. My only substantial question in general is around frequency screening and, therefore, follow-up and treatment.

I'm not aware of any new evidence that speaks directly to that issue, but it's somewhat of a more trivial or picky concern. I think the transition to an eMeasure, while not the focus of this current discussion, is fairly straightforward. So, overall, my sense is, we could endorse the discussion of evidence from our prior time and move forward.

CO-CHAIR BRISS: So, in terms of evidence, I think what we're going to try to do in these cases where we believe that the importance to measure may not have changed greatly, we'll give the Committee an opportunity

to hear that kind of a discussion and see if anybody has -- see if any of the other primary reviewers want to add to that on importance and then, see if anybody else on the Committee wants to have further discussion. Is that okay?

CO-CHAIR PINCUS: Let's see if anybody objects to moving on.

CO-CHAIR BRISS: So, the first question on the table is, is this still important to measure and report?

MEMBER SUSMAN: I think it's clearly important and there's a gap and we, unfortunately, haven't made much gains, which I sort of can trivially ask, well, jeez, if we can never improve this, why are we measuring it? But that's probably a nihilistic perspective.

MEMBER ZIMA: I just had a question. Do you have any performance gap information on teens?

DR. LUSTIG: So, actually, we're going to talk about gap next, but this vote is specifically just on the evidence base, on the

importance to measure. And so, like was alluded to, this is a case where, unless there is objection, we can accept the vote from the last time this was looked at, so this would pass and we'd move on to looking at gap.

CO-CHAIR PINCUS: But we don't actually have to vote on this specific thing.

MEMBER COLEMAN: I think I would accept it as-is. The one thing I wanted to mention was that, I think part of the evidence base and part of why it landed on moderate was due to the lack of evidence for improving depression through treatment plans of sorts, at least in adult populations in primary care.

And the one thing I wasn't sure how to reconcile that with is some of the integrated data. Because nonintegrated, mental health integrated systems, have a problem, but there's a lot of good evidence that if you integrate mental health services in the right way, you can actually get really good depression outcomes of sorts.

So, it might change that piece of evidence and it wasn't present. I don't know that it's that important. Again, I don't feel like we have to vote on it, but I just wanted to say that there's some new evidence that might fit into this in some way.

CO-CHAIR BRISS: So, I don't -- I've heard sort of a feeling around the table that this may still be important to measure and report. Anybody have an objection to that? Yes, so it sounds like Jeff thought that this might be still important to measure and report.

I don't think I've heard or seen an objection to that around the table, so could we accept the last Committee vote about this is still important and move on to the next criterion?

MEMBER COLEMAN: Yes.

MEMBER HORGAN: Do you want the other assigned reviewers to say we agree? I mean, is that the process here?

DR. LUSTIG: I think if anyone had an

objection --

MEMBER HORGAN: Yes, okay.

DR. LUSTIG: -- but like Peter said, this is -- instead of taking a vote about whether we should vote, what we're saying is, unless we hear an objection, we think we can move on.

CO-CHAIR PINCUS: Yes. Basically, what we're trying to do is, the voting takes some time and so, if there's clearly -- it's been voted on before, not a lot of new information to change that basic information, then we can just simply see if somebody objects, and then we don't have to go through the whole voting process.

CO-CHAIR BRISS: And, Connie, I think we said at the beginning that, if what you were going to say has already been said, we don't -- there may not be a strong reason to do a lot of repeating ourselves. And on this one, on this particular criterion --

MEMBER HORGAN: It's been said more eloquently than I would have said.

CO-CHAIR BRISS: Yes. I suspect we've

said what needs to be said. So, let's take, I think we'll take that as consensus for the -- that we still think this is important to measure and report and we can move on to the next criterion.

DR. LUSTIG: So, this means that we turn back to our lead discussants and we're going to move on and talk about performance gap. And so, this is where, I think, we start having some questions be raised.

MEMBER SUSMAN: I can basically say, there's still a substantial performance gap. And there's a lot of data around this in a lot of different settings. So, it seems to me that this is a measure that is still important in that dimension.

MEMBER ZIMA: I think my point was, it looks like the performance gap was based mostly on Medicare data. Any information on teens?

MS. WELLS: So, for the claims/registry version, we did only have access to Medicare data, which, of course, doesn't have many teens.

On the eMeasure, we were able to kind of look at rates in different age strata. And the 12 to 17 had an average performance rate of 53.7, which was comparable to 18 to 64. And then, we see higher rates among 65 plus on the eMeasure.

CO-CHAIR BRISS: Other thoughts on the gap?

MEMBER PINDOLIA: I just had a comment for clarification. So, I notice that every year, it's been getting worse instead of better and I was trying to understand -- and it's a pretty drastic drop from 82 to 52 percent, but then you had a higher percent of eligibles, so is it just you've got more people and it really shows the problem is worse than what we thought?

MS. WELLS: Exactly.

MEMBER PINDOLIA: Okay.

MS. WELLS: The pool of reporters went from, I think, about 1700 to 61,000 between 2012 and 2014. So, we wouldn't expect stable performance. And we think a lot of physicians are really working on the reporting mechanism at

this time, more so than the performance rate.

MS. SOMPLASKY: CMS, as more providers have moved to Alternative Payment Models or Advanced Payment Models, this is a required measure if you are in an ACO and it comes in under the claims version, under that group reporting. So, we saw a big upsurge as more and more providers started falling under that ACO umbrella.

CO-CHAIR BRISS: You see this in nearly every quality measure that I've ever looked at. As the -- in the early years, the people that were reporting on measures, as a general rule, were kind of high performers and as you get a broader pool, the average goes down.

So, it sounds like there's a thought on the table that there may be a remaining performance gap. Anybody else want to make -- do we need to vote this criterion? Yes. Anybody else have anything they want to say before we vote?

CO-CHAIR PINCUS: We don't need to

vote.

DR. LUSTIG: Sorry. The first measure always takes a while, especially with a new process. We did not have to vote on evidence, but we do vote on gap.

CO-CHAIR PINCUS: Oh, we do vote on gap?

DR. LUSTIG: Yes.

CO-CHAIR BRISS: Because the gap could have closed.

DR. LUSTIG: Right.

MS. QUINNONEZ: Are we ready to vote on gap? Okay. We are now voting on gaps for measure 3148: Preventative Care and Screening: Screening for Clinical Depression and Follow-Up Plan. I will give you your options.

For those of you who are on the phone, if you could send Kirsten your chat message, your vote over the chat line, that would be great, and we'll calculate that for you. For all those others in the room, if you would just direct your clickers this way and I will read you your

options.

For performance gap of measure 3148, Option 1 is high, Option 2 is moderate, Option 3 is low, and Option 4 is insufficient. Voting for the gaps for measure 3148: Option 1, high; Option 2, moderate; Option 3, low; and Option 4, insufficient.

We'll give you a second for all your votes to get in. Okay. All votes are in and voting is now closed. The results for performance gap of measure 3148: 74 percent voted high, 26 percent voted moderate, zero percent for low, and zero percent for insufficient. So, this passes with 74 percent.

CO-CHAIR BRISS: So, I think that moves us to our reliability discussion, is that right? So, anybody want to tee up reliability of the measure for us? So, let's let somebody other than Jeff take a turn. Maybe, David, you want to -- can you tee up reliability for us?

MEMBER EINZIG: So, I don't know if I'm the best person to speak on reliability and

validity, I'm not the statistician. If any other takers want to do it? I certainly can otherwise. All right.

CO-CHAIR BRISS: Why don't you take a first shot and --

MEMBER EINZIG: Okay.

CO-CHAIR BRISS: -- we'll go around the table if we need to?

MEMBER EINZIG: So, I'm just going to go with using the guidance from the -- going off comments. The preliminary rating for reliability was high from the Committee. I really have nothing to add to that, other than just that.

CO-CHAIR BRISS: Anybody like to add?

MEMBER SUSMAN: Ditto.

(Laughter.)

CO-CHAIR BRISS: Yes, Harold?

CO-CHAIR PINCUS: So, I had a question about the reliability for the follow-up information, because it seems to me that that is potentially more problematic. And so, what do we know about that right now?

MS. WELLS: Well, I think the follow-up can be a feasibility issue, in terms of collecting it. Although that affects the performance score and even -- so the performance scores themselves we did find to be highly reliable in that the ratio of signal to noise was high. And that was true for even clinicians with fewer than ten patients.

MS. SOMPLASKY: For the claims and registry version that we're currently discussing, there are HCPCS codes that allow them to document the follow-up plan. And it says that they either have or they have not. Much easier, when we get to the eMeasure, being able to have that in a reported field. But on the claims and registry, it was much more reliable.

MEMBER SUSMAN: Isn't the big issue, to me is, measuring a plan versus measuring what actually happens and persistence of treatment? Which presumes that the data is available. So, in some ways, I think the bar is low enough that there's probably pretty good reliability here.

CO-CHAIR PINCUS: Is there a concern that the bar is too low?

MEMBER SUSMAN: -- for the bar, so I'm not sure putting it higher will make it easier. Only if you put it high enough that we can walk under the bar.

CO-CHAIR PINCUS: Lisa?

MEMBER SHEA: I was just interested in one of the exclusions, of people who refuse, and how often that's used, because it seems to me people who are depressed might be more inclined to not want to engage in any kind of activity.

MS. WELLS: Right. So, people who are actively diagnosed with depression are actually excluded from the measure. I think the assumption is that they're already in treatment, so they don't need to be screened again.

(Off microphone comment.)

MS. WELLS: Well, I know in the eMeasure, we actually see very low patient refusals. Let me check on the claims/registry. I believe our exclusion rate was -- I think it

was fairly low for patient refused. We're seeing most of it coming out of the active diagnosis of depression.

And that's a great question, because a patient refusal can be -- it's kind of -- it's got its pluses and minuses in a measure. I think our perspective is that we're concerned about clinician/patient tension if patients refuse the measure and then that affects the clinician's score. So, we do keep that in there, but we actually don't see it used very often.

MEMBER LARDIERI: Yes, and just a question. So, how do those G-codes get entered? And providers, and I'm concerned about the ones where they say they didn't do something, they actually enter that, and where?

MS. SOMPLASKY: That is part of -- so, within folks who may -- we have many folks who have an EHR and are still reporting that registry version, and that just gets into a whole different set of issues with vendors' ability to report these measures.

But those HCPCS codes are loaded. For folks who really still are on paper, those HCPCS codes are part of their super-bill, if you will, that they are picking when they report the measure.

And you do, for PQRS, report your intention to be reporting certain measures. That's something that you are setting up as part of your submission, if you will.

MEMBER LARDIERI: And when PQRS goes away?

MS. SOMPLASKY: It's now -- as I said, it will be in the MIPS program. The claims and registry version will remain in case those providers not included in this new QPP program, such as social workers, psychologists, physical therapists, folks who have traditionally been reporting this are able to still report it, it's just not going into a payment program.

MEMBER JENSEN: I have a question about exclusions. So, patients are excluded if they come in for emergent visit, which makes sense to

me. But what if that emergent visit is a result of some risk-taking behavior that is related to depression? Do you know what I'm saying, does that make sense?

MS. WELLS: Yes. And, actually, the medication exception allows providers to still screen those patients and include them in the denominator and numerator if a depression screening is appropriate.

CO-CHAIR BRISS: And it might help for all the exclusions just to have a sense of how important that is, it sort of relates to how often it's taken. So, are any of the exclusions common?

MS. WELLS: The active diagnosis of depression is by far the most common --

CO-CHAIR BRISS: Okay.

MS. WELLS: -- exclusion that we see, yes.

MEMBER MAZON JEFFERS: I had just a question back to is the bar too low point that was raised earlier. And it's hard for me to

think about this in a vacuum, are there other measures that are looking a little bit more closely at the quality of the plan that's in place and whether or not that plan has been acted upon? Could you just put it in context for me?

MS. SOMPLASKY: We are the only measure that actually asks for a follow-up plan. And that's been the differentiator. There are a few measures that have not sought to get their re-endorsement for NQF that were doing screenings, but they are no longer NQF measures.

But we're the only one who is looking for and asking for documentation of a plan. We then -- we don't get into being prescriptive, though, what that plan has to be, because we don't want to scare people off from doing that basic important screening.

MEMBER COLEMAN: I just wanted to mention that, I guess, I think it's important to also screen those with depression and bipolar. And, in fact, we know primary care has challenges in treating those folks, so I'm not sure why you

would exclude them.

In my mind, the prevalence is going to be way higher in that population, of sorts, and the assumption that they're getting treatment is probably not a good one in primary care in general. So, I just think it would be great if you would think about including them.

MS. SOMPLASKY: The focus when this measure was initially developed, CMS wanted to see screenings done. We were not looking to treat, because, again, we don't want to scare off those folks who otherwise feel like, wait a minute, now I'm going to be responsible for treating that depression. We're trying to identify new cases and make sure that they get the follow-up that they need.

MEMBER COLEMAN: I guess with the -- but you are following to see if there's a treatment plan, so I think anybody who's suffering a current period or episode of depression should have some sort of treatment plan. And I just think it's not the best

assumption to assume that because they have a diagnosis, they are in treatment or have a treatment plan. If that makes any sense.

MEMBER PATING: Just wanted to ask a question between this and other measures, I don't know if that should be now or in any kind of harmonization discussion. But the NCQA depression measure, I can't remember, first of all, whether we have reviewed that here and whether that is a competing, you consider that a competing measure.

MS. WELLS: So, we talked to NCQA and they let us know that they actually didn't submit their depression measures for re-endorsement. So, I actually don't think those are in use currently, that's my understanding.

MEMBER PATING: Okay.

CO-CHAIR PINCUS: Yes. I just want to endorse what Shane was saying. Actually, a number of years ago, Jurgen Unutzer, Wayne Katon, and I did a study looking at claims, from a large group of Medicare patients, looking at claims

compared to PHQ-9 scores and found that, if you were looking for the people who were really depressed, it was the people that were currently in treatment.

CO-CHAIR BRISS: Raquel?

MEMBER MAZON JEFFERS: Does the -- if someone has a -- if someone is receiving a prescription for a psychopharmaceutical for depression, does that include -- is that counted as a treatment plan?

MS. SOMPLASKY: They would be excluded from the denominator, because they would already be considered to have a diagnosis of depression.

MS. WELLS: If the pharmaceutical was in response to the screening, they would be captured as a -- that would count as a follow-up.

MS. SOMPLASKY: With a new.

MS. WELLS: Correct.

DR. LUSTIG: If I could just jump in to make a clarification, measure 0518, actually they've withdrawn from consideration for endorsement. We do still have some measures in

our pediatric and child portfolios, but that measure is no longer related or competing.

CO-CHAIR BRISS: So, does anybody have anything that hasn't been said that they'd like to get on the table before we vote on reliability? Hearing none, let's try to vote.

MS. QUINNONEZ: Voting is now open for the reliability of measure 3148: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient. We are now voting on the reliability of measure 3148: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

All votes are in and voting is now closed. For the reliability of measure 3148: 35 percent voted high, 61 percent voted moderate, four percent voted low, and zero percent voted for insufficient.

CO-CHAIR BRISS: So, that passes and we can move on to validity. And maybe, Shane, would you be willing to speak first on this point?

MEMBER COLEMAN: Sure. Which, sorry,

which one are we on again?

CO-CHAIR BRISS: Validity.

MEMBER COLEMAN: So, I think we've already talked about some of this. I think the validity data looked pretty good. I was a little bit confused about some of the G-codes that were included in the numerator.

I thought that it was saying that there was no documentation that the screening had occurred and it was still be counted, I thought, which I wasn't sure about. But, overall, I actually thought the data looked good. Let me look at my other comments.

MS. WELLS: So, if I could clarify your question about the G-codes. There's sort of two rates, there's the reporting rate and the performance rate. And so, the four G-codes that are showing in the numerator are actually for the reporting rate.

And so, that includes patients who both met the numerator and did not. But for the performance rate, if there was no screening or if

there was a screening without follow-up, that actually is not in the numerator.

MEMBER COLEMAN: I think I probably gave my other comment a little bit early. That is, including those with the diagnosis of depression or bipolar, maybe speaks more to validity than reliability, so I guess I'd kind of move that into this segment.

CO-CHAIR BRISS: Would any of the other primary discussants -- Connie?

MEMBER HORGAN: I agree with the previous comments. I'd like to question the new testing that was done on the face validity with the 12 clinicians. That seems like a small number, could you talk a little bit about the noes or the disagree? I think it was about 25 percent?

MS. WELLS: Sure. So, the noes mentioned things related to documentation burden. One wanted more specificity around the instruments that we use, but we actually wanted to be open-ended so that physicians could use, as

long as they were valid instruments, that they could use the ones that they were comfortable with. And I believe those were the major concerns that were raised.

MEMBER SUSMAN: I think my biggest concern about this is just around frequency. And I wonder if developers have investigated this at all, or in your review of the literature, around screening each visit versus some specified period of time. Recognizing that depression is a waxing and waning phenomenon.

MS. SOMPLASKY: We have not found any evidence with a recommended frequency. We do not require every visit, we require it to be once in that screening year.

They can screen more and we look at the -- if something has changed, if a patient has come in, for example, with weight loss and that physician feels like maybe we ought -- even though six months ago, you were okay, let's screen again. But we only ask for it once in that reporting period.

MEMBER SUSMAN: And that's fine. I'm not -- I don't think this is a major issue from my perspective. I don't know about --

CO-CHAIR BRISS: And it looks like, Les, is your card up?

MEMBER ZUN: Let me try this again. I'm a little concerned that there is no empiric validity data for this measure. I just -- why is there reliance on the face validity?

MS. WELLS: There's a few reasons. We weren't able to find, like, a comparable measure with the same population to do a more empirical validity testing, we did look into that. We would have liked to have included data element validity, unfortunately, we had a very constrained timeline. Just to be very candid with you, we were unfortunately unable to get that data together in time.

CO-CHAIR BRISS: And is there more that should be said about threats to validity? We've talked about exclusions already at some length. Any inappropriate exclusions and anything else

anybody wants to talk about?

MEMBER PARISH: The one exclusion that hasn't been mentioned would be adjustment disorder with depressed mood, where you would have a treatment plan that may be overly aggressive when it's an adjustment disorder versus depression.

CO-CHAIR BRISS: So, are you arguing that there might be an -- you might want to add an exclusion? Is that what you're -- is that the argument you're making?

MEMBER PARISH: Yes.

MS. WELLS: An exclusion for, I'm sorry, did you say adjustment disorder?

MEMBER PARISH: With depressed mood.

MS. WELLS: Oh, okay. Thank you for that, we'll look into it.

CO-CHAIR PINCUS: And just a little bit more about the rationale for why that should be excluded?

MEMBER PARISH: If someone comes in with an adjustment disorder with depressed mood,

to have a requirement, grant it with a treatment plan that requires referral to a licensed counselor or psychologist/psychiatrist or be put on medication, that might be overly aggressive when they may need more of a pastoral counselor or other social supports.

MS. SOMPLASKY: We are not prescriptive as to what that follow-up plan is. So, if they referred to a pastoral counselor, that meets the intent of the measure. That was just an example, about referred to a psychiatrist or psychologist or medication. A follow-up plan could also be, have the patient return to clinic in two weeks to do another recheck and perhaps rescreen.

CO-CHAIR PINCUS: So, which is basically watchful waiting?

MS. SOMPLASKY: As long as the documentation is there.

DR. LUSTIG: And I just want to clarify that, in terms of NQF process, we don't require score-level validity and we don't require updated testing if it has passed. I mean, we can still

discuss it, but those are what our criteria are.

MEMBER SUSMAN: When you say, score-level validity, are you talking about elements within or the total score? I'm trying to -- I'm a little lost.

DR. LUSTIG: I would do my best, but I'm going to rely on my statistician over here.

MS. JOHNSON: So, the data element validity is things like checking inter-rater reliability --

MEMBER SUSMAN: Right.

MS. JOHNSON: -- looking at the individual data elements and comparing, make sure you can consistently pull those. Score-level validity, that's the NQF-speak for looking at the results and comparing results.

So, often, you might see, looking at things like how does this measure track with other measures, if they are available? And sometimes it's hard to find measures that you can correlate with.

MEMBER SUSMAN: Okay.

MS. JOHNSON: That sort of thing.

MEMBER SUSMAN: Thank you.

CO-CHAIR PINCUS: Tami?

MEMBER MARK: So, what is the validity requirement? I mean, you're essentially saying there's no need to recheck validity for the maintenance measures?

MS. JOHNSON: We do not require, at this point, that developers retest. It would be great if they did, but we do understand there are resource constraints that may prohibit that.

CO-CHAIR BRISS: Rhonda?

MEMBER ROBINSON BEALE: I have a question under exclusions, I'm not quite sure I understand. Under exclusion, you said situations where the patient's functional capacity, I understand that one, or motivation to improve may impact the accuracy of results in the standardized depression assessment tool. Question, give me an example of a patient with motivation to improve may impact the accuracy of the results?

MS. SOMPLASKY: We have had -- one of the discussions that came up with our expert work group were patients who were being forced to be examined. We've had juvenile cases where they were asked to be evaluated and there's noncompliance, they don't want to be part of the screening.

Or we also had, with certain prison situations, where the patients were saying that they were depressed and there were questions about it. So, it was an exclusion for that provider to be able to make that judgment.

MEMBER ROBINSON BEALE: So, the reason why I ask this is that the -- in health plans, you look for ways of understanding whether or not the primary care practice is screening sufficiently enough. This exclusion is kind of unclear. If I was a practitioner and didn't ask that question, I really wouldn't have known what that meant.

My concern is that, as a plan, I would use this measure to understand whether or not

enough screening is going on. With the provider being able to exclude certain people out of that denominator, I'm not sure it's giving me a good reading on the completeness of their applying the screening for depression in a population 12 years and older. Does that make sense?

MS. WELLS: It does, and thank you for that. Again, I think this is one of those things in measurement where there's pluses and minuses. And we didn't want physicians to be held accountable for screening people who are court-mandated to be screened and where the treatment wasn't likely to yield much. So, we give them the option to take them out or keep them in, but I do appreciate your point.

CO-CHAIR BRISS: One thing that I think might help you as you think about submissions in the future is giving us a sense of, perhaps a better sense of how many total people are being excluded. Sorry. One thing that might help you in future submissions is giving us a better sense of how many total people are being excluded.

MS. WELLS: Yes. And we actually didn't see -- I don't know that we saw anybody excluded for that particular. We saw some with the patient refusal or emergent medical, although I don't think we actually observed any in that particular exception.

MEMBER ROBINSON BEALE: I think, if people are using this for accountability, I think that would be helpful and that's a good suggestion, Peter, to be able to collect that kind of information.

CO-CHAIR PINCUS: One question about clarification and the second one just in terms of validity. So, in terms of clarification, this applies not just to primary care providers, it applies to all specialists, including behavioral health specialists. Just to clarify that.

Number two, in terms of the validity data, it seems to me, since this is a registry or claims-based data, the important thing is to see whether what people sort of put down on a claim has actually happened and whether it's documented

in the chart. And it looks like there was reasonable kappas as you did that.

The question I had had to do with the representativeness of the population of physicians that you looked at. And it comes up because I recently reviewed a paper that actually showed very high performance on this and other ratings, but it turned out to be -- it was particularly large, highly infrastructured primary care organizations.

And I just want to get a sense of what kind of practices were looked at in terms of the physicians that were included in the validity study.

MS. SOMPLASKY: For the claims and registry, we do a random sampling. Now, it's voluntary for those practices to submit the copy of their charts. We do reimburse for the copying, but, again, we can't make it mandatory for them to send it in. So, I don't know, we just do a representative of the claims and --

CO-CHAIR PINCUS: So, all --

MS. SOMPLASKY: Yes.

CO-CHAIR BRISS: So, Raquel? I'd like to move us pretty quickly towards a vote on this criterion. So, Raquel?

MEMBER MAZON JEFFERS: Yes. So, exactly the question you just brought up, this measure is not only for use in a primary care setting, but also in a behavioral health, like specialty care setting regular outpatient program.

So, I have trouble getting my head around the first part of a screening in an outpatient setting, because someone has showed up for an outpatient visit in a specialty care setting, I don't believe they need to be screened. What you're really testing is whether they have a plan in place, is that how you see this measure applying in that setting?

MS. SOMPLASKY: No. It's open to -- there are people who maybe don't have that diagnosis of depression and are being seen by behavioral specialists or psychologists and

social workers, who can also report this, physical therapists, occupational therapists. So, it's just part of that assessment of that patient, the overall assessment.

MEMBER MAZON JEFFERS: But a screening is different than a diagnosis.

MS. SOMPLASKY: Correct. And that's what we're saying is that we are trying to increase the screening rates to get to a point of diagnosing that depression.

CO-CHAIR BRISS: All right. So, I see no further cards up. I'd like to try to move us towards a vote on validity and see if we can get through another vote.

MS. QUINNONEZ: Voting is now open for the validity of measure 3148: Option 1 is moderate; option 2 is low; and option 3, insufficient. Option 1, moderate; option 2, low; and option 3, insufficient.

DR. LUSTIG: And I just should have jumped in to say, this is one of those cases where when you follow our algorithm, the highest

possible rating is moderate, because of the lack of score-level testing.

MS. QUINNONEZ: Okay. All votes are in and voting is now closed. For the validity of measure 3148: 78 percent voted moderate, 13 percent voted low, nine percent voted insufficient. So, this passes with moderate.

CO-CHAIR BRISS: So, we're getting close to the final turn.

(Laughter.)

CO-CHAIR BRISS: The feasibility to measure, how about, Andrew Sperling, can you tee this up for us? Or did I do this wrong? I'm sorry, Andrew, I called on the wrong person. No wonder I was getting that look. Connie, can you tee this up for us?

MEMBER HORGAN: It's highly feasible on everything that was noted.

CO-CHAIR BRISS: All right. Is --

MEMBER HORGAN: Do you want me to go through it?

CO-CHAIR BRISS: So, adequate

feasibility, if you think it's straightforward, there's nothing wrong with being straightforward. Shane, wanted to add something?

MEMBER HORGAN: They only provided the feasibility on electronic sources, they did not report any implementation challenges at all.

CO-CHAIR BRISS: Shane?

MEMBER COLEMAN: I think my question around feasibility was just around the treatment planning piece, which corresponds to each of those pieces that were acceptable as the treatment plan and how does it kind of fall out across systems, I guess, in a standardized way?

MS. WELLS: Sure. So, with the claims/registry version, they do make use of the HCPCS codes for reporting. And so, they select one of six and that would decide whether it's excluded or meets or fails performance. And so, that's why the feasibility of the claims/registry is pretty straightforward.

CO-CHAIR BRISS: Anybody have anything to add? Why don't we try to vote on feasibility,

please?

MS. QUINNONEZ: Voting is now open for the feasibility of measure 3148: Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient. For the feasibility of measure 3148: Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

Looking for one more vote, if you could direct your clickers this way. Okay. Voting is now closed. For the feasibility of measure 3148, the results are: 55 percent voted high, 41 percent voted moderate, five percent voted low, and zero percent for insufficient. This passes.

CO-CHAIR BRISS: So, that gets us to the last one, usability and use. Maybe, Connie, can I call on you again to do the last -- to bring us home? Usability and use?

MEMBER HORGAN: I don't -- oh, yes. The usability, it was widely used in various CMS programs. It will replaced by MIPS and has been --

CO-CHAIR BRISS: Closer.

MEMBER HORGAN: Closer?

MEMBER SUSMAN: You got to get right --

MEMBER HORGAN: Oh, I'm sorry. It is widely used in various CMS programs. And the only concern I have, which was raised earlier, was the performance declining over time from 2011 to 2014, from 82 percent to 52 percent, but it was explained by the number of professionals increasing. It still is not widely used, so we are sort of angels dancing on the head of a pin when you have such small numbers. That was a highly statistical statement, I think.

(Laughter.)

CO-CHAIR BRISS: Anybody else want to add to that? David?

MEMBER EINZIG: So, just in terms of leading to improved outcomes, it would be nice to see more data on, if this is used, does it actually lead to improve outcomes? Do people actually follow through with the treatment plan and does it help?

And then, on another level, and this is sort of my soapbox, but I think there's a lack of treating in primary care, where they can screen for things, but then as far as actually treating, I think more focus should be put on earlier levels of folks in medical school and nursing school and residency programs and look more at piece also.

MEMBER PATING: My question is on the usability of the measure. With the declining averages, how does that affect reporting and percentiles over time? I could see you can still use it for quality improvement internally, but for external comparisons, I just was wondering, what's been the impact of that, when you're comparing across plans?

MS. SOMPLASKY: We don't compare across plans. We -- the only claims that we have access to and look at are the CMS Medicare claims. So, we don't have anything across other plans, we don't have access to that.

MS. WELLS: Just to add to that, I

think what might be happening with the performance rate, again, is that there's a lot of new participants in the program that are really focused on their reporting mechanism, and it might be part of an annual wellness visit menu of screenings, that kind of thing.

So, we think that they're focused on reporting more than performance right now, because I don't believe there are penalties on this data on their pay just yet. So, hopefully that would -- that allows them to kind of get the process in place before those penalties get -

MEMBER PATING: I think that's really my concern. So, if this ever gets incorporated into any sort of Stars measure and the n keeps changing or the denominator keeps changing and driving down the average, you just get really skewed performance year-over-year. So, I would hope the measure gets more stability before you develop that as a publicly accountable measure.

MEMBER ROBINSON BEALE: I think David kind of asked the question I had, and that was,

the usability of this measure as a comparative measure, and I think you've answered that that was not the way it was designed at this point. Then, let me understand, what was the purpose of the measure in terms of how you are using it? In terms of your organization?

DR. GREEN: I can take this one. This is Dan Green from CMS. So, we're using the measure to identify, obviously, the under-diagnosis of depression and encourage primary care, or any physician or caregiver that wants to participate in the measure, to screen their patients for depression. I think the declining numbers with more people reporting it is good evidence of a gap and an opportunity for improvement.

In terms of publicly reporting these, I mean, any measure that are in our quality programs, PQRS currently or PQRS through 2016 and MIPS going forward, could be reported publicly. That's not to say that we will report every single measure, but any measure that's not in its

first year of use could be reported.

As far as payments, we look at the benchmark for the measure and depending on whether someone is above or below the benchmark would determine whether or not they receive actually an incentive versus a neutral adjustment or a payment adjustment, based on that particular measure. But, obviously, that's only one small part in the whole MIPS program scoring.

MEMBER ROBINSON BEALE: So, thank you for that explanation. So, just a question, in the way that it was field tested, were providers, did they know or was there an incentive or disincentive attached to this process at the time it was being tested?

MS. SOMPLASKY: No, we --

DR. GREEN: What year did we do that, guys?

MS. SOMPLASKY: We tested it last year. Providers did not receive compensation for participating and they didn't know we were collecting their data --

MEMBER ROBINSON BEALE: Okay.

MS. SOMPLASKY: -- for, obviously -- that was a claims data analysis that was done. But those that did participate and provided us with an interview and feedback were given a small incentive for participation in that, because we were onsite, we did work -- and that's as we get to the eCQM, but that involved workflow analysis, provider time, staff time.

MEMBER ROBINSON BEALE: Okay.

DR. GREEN: And there was potential compensation from the PQRS and value modifier program, but not specifically -- it wasn't identified that this particular measure would get you in the positive or negative adjustment. It was just one of the measures as part of the requirement that you could report as part of the requirements for the programs.

MEMBER ROBINSON BEALE: And the benchmarking that you mentioned, that was based on your organization and those who were field tested?

DR. GREEN: So, the benchmarking is particularly important for the MIPS program and it is based on historical data that we have for the measure. Or if there are national standards, I believe they will be factored in, but I'm not 100 percent on that.

MEMBER ROBINSON BEALE: Thank you.

DR. GREEN: They're certainly based on our information.

CO-CHAIR BRISS: So, we're -- as sometimes happens on the first one, we're a little bit behind. I have three more cards up, if we could go quickly through those three people and then, let's try to move to a vote. So, Tami?

MEMBER MARK: Yes. Just to follow up on David Einzig and Brooke Parish's point about screening and the relationship to good treatment, is there any thought about harmonizing this measure with the PHQ-9 depression measure?

MS. SOMPLASKY: We've had discussions with our expert work group about it, but right now, they feel that our piece with the screen and

because we're not prescriptive about it being the PHQ-9 that is used, we've not had any further discussions about it. But it is part of the discussions that we have with our expert work group.

CO-CHAIR BRISS: Shane? Or maybe, no, I'm sorry, Les?

MEMBER ZUN: So, I have one concern. I think, conceptually I think this is a great measure, but I'm reading the one denominator exclusion that is problematic, patients in an urgent or emergent situation where time is of the essence, blah, blah, blah.

Well, emergency departments are quite a unique place, where the patient may wait in the waiting room for a long time, but this is a very problematic burden, because if I identify someone who is depressed, then what am I doing there? Do I have the resources to provide that patient, if they screened in? And there are so many other screens that we have to do in emergency medicine, that I --

MS. SOMPLASKY: This is not for emergency medicine, this is outpatient codes only.

MEMBER ZUN: Well --

MS. SOMPLASKY: So, we --

MEMBER ZUN: -- if it's an outpatient code, that would be emergency medicine codes, because they're all outpatient codes.

MS. SOMPLASKY: And they -- I'm sorry, I should be clearer --

DR. GREEN: I think she means E/M codes.

MS. SOMPLASKY: The evaluation and management codes for emergency medicine are excluded from this measure.

MEMBER ZUN: Okay. Thank you.

CO-CHAIR BRISS: You get the final word.

MEMBER SHEA: Yes, thank you. I was just wondering, in terms of, I know as people are reporting in, that 7.5 percent of the eligible professionals reported in, how does that compare

with other measures? Are they not choosing this measure as much as others or -- just maybe as an indicator of the usability of it.

MS. WELLS: I think it's in the top 20, maybe, it's definitely not, like, in the top five, but we're seeing it get reported a lot more, I think because it aligns with a lot of programs out there with ACOs and so forth.

CO-CHAIR BRISS: So, with that, let's try to move to vote on usability and use.

MS. QUINNONEZ: Voting is now open for the usability and use of measure 3148: Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient information. We're voting on usability and use of measure 3148: Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient information.

Yes, it will capture your last vote. Voting is now closed. For the usability and use of measure 3148: 14 percent voted high, 77 percent voted moderate, nine percent voted for low, and zero percent for insufficient

information. This passes with a moderate.

CO-CHAIR BRISS: All right. So, this measure, we don't actually have to talk about related and competing measures, so I think we've successfully completed our first one.

DR. LUSTIG: Actually, the one that --

CO-CHAIR BRISS: But we have an overall --

DR. LUSTIG: -- we had on the PA has been withdrawn for consideration, so there's no longer a related measure to discuss. Measure 0518.

CO-CHAIR BRISS: And do we have an overall vote on it?

DR. LUSTIG: So, we have to do an overall vote.

CO-CHAIR BRISS: Okay. So, one more set. Anybody have final words before we do an overall vote on the measure? So, let's vote one more time.

CO-CHAIR PINCUS: Actually, I just had one comment, just in terms of the related and

competing. At some point, we may want to look over the range of depression measures and look at the span of them and their interrelationship. Not right now, but I think at some point, I think it would be a good idea.

CO-CHAIR BRISS: David?

MEMBER PATING: I'm thinking specifically the Minnesota Community Measurement, NQF 0711 and 1884. Those are the ones that I was referring to earlier with NCQA.

MEMBER GROSS: And, Charlie Gross, with Anthem. Just to follow up, Harold, to your point, in terms of how we vote yes/no, I mean, I'm going to vote yes on this, it's useful, but the competing measures from an operating perspective and a provider prescriptive is overwhelming.

So, while these things are very useful, the reality is, if multiple measures are out there for the same dimension, many providers, sort of learned helplessness, wonder, which ones should I use, and sometimes the default is, I

don't use any of them.

So, I think the, not the developers of the measure, but I think NQF needs to wrestle with that issue, because it's a huge one, both from the provider perspective and the payer perspective, when we go out and say, which of these alternative measures should we be rolling into our value-based purchasing agreements?

DR. GREEN: Peter, can I just comment real quick on that?

CO-CHAIR BRISS: Yes, Dan.

DR. GREEN: Thank you. So, I think that's an excellent point. And I'm sure as you know, we are part of the AHIP group that is looking to have a harmonized or adopt particular measures that all the insurance companies can agree on. So, I'm not sure that we've hit mental health yet, but I think that's a great suggestion. I'm sure it will be on one of the panels in the future. Thank you.

CO-CHAIR BRISS: Anybody else have any urgent things that they want to say before we do

an overall vote? Seeing no cards, so, this one is simpler, yes or no.

MS. QUINNONEZ: Voting is now open for the overall suitability for endorsement of measure 3148: Option 1, yes; option 2, no. We are voting on the overall suitability for endorsement of measure 3148: Option 1, yes; option 2, no.

All votes are in and voting is now closed. For the overall suitability for endorsement of measure 3148: 100 percent voted Yes.

CO-CHAIR BRISS: I don't know that I've ever been in a 100 percent.

DR. LUSTIG: So, earlier, people were asking me how we need to consider the eMeasures differently from the maintenance measure and so, I'm glad that I don't have to give the talk, I have my colleague Jason Goldwater here, who's going to explain everything you need to know.

MR. GOLDWATER: No pressure at all, Tracy, thank you. So, good morning, everybody.

Always enjoy this part of the meeting. So, the wonderful world of eMeasures. I'm sure all of you have been looking forward to this since the moment you stepped in here this morning, because nothing gets people's blood roiling than electronic clinical quality measures.

So, I am not here to explain the universe of this, I certainly could and if you really want to get that geeky, I'm more than happy to do so. But what I really want to do is sort of talk about eMeasures from a larger, sort of 50,000 foot view, of sort of what eMeasures are, the different types of eMeasures that come into NQF, what we look for when we see eMeasures, and what you can expect.

And then, when that's done, I'll happily answer whatever questions you might have. I'm going to hope that this is going to work. Oh, good, okay. So, eMeasures are not something that's new. I think all of you who have been doing this long enough probably know that.

eMeasures were first brought up by

CMS, back in the good old days when they were HCFA, do we all remember that? Yes, that was the long -- we're all aging ourselves here. And they wanted to, in parallel with then George W. Bush's Directive to have an EHR in every hospital and physician office by 2014, and which that did not actually happen, but we did come fairly close, that if there was a way of electronically recording information, there should be a way of electronically reporting it.

Because the measure process at that point was abstracting charts manually, which was not the most efficient way of doing so. And so, they began with the Doctors' Office Quality-Information Technology Project, or DOQ-IT for short, any of you remember that? Right.

So, I was the Project Manager for that, which I rarely, if ever, admit to people. But that was the first attempt. And DOQ-IT had the greatest of intentions, but it did not necessarily work, because EHR adoption at that point was under 20 percent nationally.

So, now, we're sitting here in 2017, where it's 80 percent, and so, there is a greater ability now to be able to record data and report it out electronically. And when measures come in that are electronic to NQF, they usually come in in one of four ways.

So, the first is a de novo measure, which, as you know, is a brand new measure. Rather than it be specified through a claims extraction process, it is specified specifically as an electronic measure and it has to adhere to the same measure submission and testing process that any NQF measure would have to submit to.

Then there are what we call respecified eCQMs. And so, those are measures, and I believe that's what you're going to see, so, that's a measure that was a chart-abstracted measure and has been used, but now it's up for maintenance and so they've respecified that.

So, they've taken the chart-abstracted measure and they've made it into an electronic one. And I'll tell you the components later

about what goes into that. And so, it must adhere to NQF's measure submission and testing process.

Then there are what we call legacy electronic clinical quality measures. So, those are chart-abstracted measures that are currently used in national federal reporting programs, like PQRS or the IQF program or maybe even be recommended for MIPS.

But they are chart-abstracted and the desire of CMS, and really of the federal government, is to move those into an electronic form. And so, a legacy measure, the only thing that differentiates that between a respecified measure is the way the measure is tested, because they could use something called the Bonnie tool, which is owned and operated by MITRE. Before any of you ask, because I get this all the time, what does Bonnie stand for? Nothing.

(Laughter.)

MR. GOLDWATER: I have no idea what it stands for, if I had to say, it is probably the

name of the developer's pet or child. I mean, I used to be a developer and we're not creative folks, we name it after the first available thing that we see.

(Laughter.)

MR. GOLDWATER: So, that's probably what it was. We do not have -- there are no legacy measures that are coming under consideration to all of you. And we are slowly sunsetting that, because there were a lot of legacy measures over the last couple of years, but that is gradually winding down.

And then, lastly, there is what we call the Trial Approval Program. You will not have to worry about that, but just for your own education, those are measures that are innovative and new and there's clearly a need for, but it's difficult for them to meet the NQF criteria for endorsement, namely, in the testing side.

But the measure is deemed to be so important, it would be a waste to not actually put the measure forward if it could actually

benefit quality because of those testing restrictions.

So, trial use is, the measure gets through the same review as any NQF measure does, but it doesn't get endorsed. What you're then doing is, not endorsing it, but you're accepting it into the trial use program and it actually gets put into the field and it can be used and it collects data while it's being used.

And once the developer feels they have enough data, they pull the measure out of the program, they analyze the results, and then, they bring it to you again, this time for endorsement. And we have roughly a dozen measures in the trial use program at the moment.

But, again, that's not applicable here. I almost said next slide, I really got to get out of that. All right. So, what are the specifications for an eMeasure? So, the first one is what we call HQMF, or the Health Quality Measure Format.

So, what on Earth is that? So, I'm

sure all of -- how many of you use the internet for shopping? And don't be shy, right? All of you do? Right. I mean, why do we go to the store anymore? I mean, I found Boxed -- do you know what Boxed is?

Boxed is like going to Costco without actually having to go to Costco, you can order all of these big items and they're shipped right to your door and you don't have to deal with Costco. It's the most wonderful thing in the world, I will never go to Costco again.

(Laughter.)

MR. GOLDWATER: And the way -- it's true, why would you want to stay at Costco? So --

DR. LUSTIG: They give out free samples.

MR. GOLDWATER: What's that? They do.

(Laughter.)

MR. GOLDWATER: It's a great idea. Anyway. The way that that works, the way any online shopping works is, you input data and your

credit card number, any personally identifiable information that's deemed necessary, and you send it to the website that is going to be receiving it for your order.

And the way that that is coded is usually what we call the Extensible Markup Language, which has been around really since the early 1990s. And all that is is a very basic common way of tagging and organizing the information, so that it goes from your system to the receiving system, the information is interpreted, it's recorded, and your order is placed correctly.

This is the same thing, because the measure has to go from your system, has to be inputted into a system, and it has to be able to produce the result. So, HQMF is basically a standard style way of organizing an electronic measure.

It basically is the way it needs to be laid out, what the tags need to be, what the meaning of the information is, so that it can go

into any EHR system, regardless of what the vendor is.

The way developers tend to create electronic measures is the Measure Authoring Tool. And in that, when you create a measure from that, it comes out in the correct format. The MAT is not required to do HQMF, you're welcome to do it by hand, I don't know why you would do that, but there are people that do.

But we do require that it has to be in the Health Quality Measures Format, because it's a very specified style and it ensures that it can be used, then, in any EHR.

It also has to have value sets. So, what are value sets? So, value sets are really the way you represent clinical content. So, major depressive disorder would be represented as a value set and it would come from a controlled clinical vocabulary.

And there are a variety of them. There are ICD codes, which a lot of you are probably very familiar with. There's SNOMED, are

people familiar with what SNOMED is? All right. That's a very rich, very robust, clinical vocabulary that covers numerous, numerous clinical concepts.

There's LOINC, which is used for laboratory orders and tests. There's RxNorm, which is used for medications. And these are standards that have been promulgated by the Office of the National Coordinator and are really commonly used throughout the community.

But value sets have to be part of a measure. And value sets are maintained by the National Library of Medicine in their Value Set Authority Center. So, any eMeasure that comes to us has to have value sets that are in the VSAC, as we call it for short, and they have to be published.

So, they cannot just be some random value set that a developer created that nobody else can see, it has to be a national value set that is published in the VSAC. If the eMeasure that comes in, and I look at all of these, it's a

joy, trust me, especially when you get 90 of these, but the measure developer -- if I don't find that there is a value set that's not published, if I can't find it in the VSAC, then I have to call the developer up and say, why did you use this?

And they'll give me a reason and I'll say, that's great, go into the VSAC and find the code that matches what you have. If you cannot find it, then we'll have to work on getting that, what you've created, published. Which we can do, but you have to have a nationally published value set.

And the reason for that is, because when you have a national one, it can be reused by measure developers everywhere and it ensures consistency. Because one of the problems when eMeasures were first created is that everybody was creating value sets and there was no centralized way of keeping it.

So, that means ten developers could create ten different value sets for generalized

anxiety disorder. It means the same thing -- I mean, I'm not the expert, I'm just assuming that it means the same thing, but everybody was coding it differently.

So, to try to harmonize that, we just want one value set for that particular condition. If the published value set does not exist, the measure developer must demonstrate that their value set is in draft form and then they have to wait to publish it to the VSAC.

So, the Value Set Authority Center, which is like a library of value sets that the National Library of Medicine has. And they can go in there, submit it, and get it published. It's not a difficult process whatsoever and we usually do provide guidance, if necessary.

CO-CHAIR PINCUS: So, just to be clear, so, for example, the generalized anxiety disorder example, so, would ICD-10-CM be a value set?

MR. GOLDWATER: It could be, absolutely. ICD, CPT, HCPCS, SNOMED, LOINC, we can go on and on. There's a lot of them. What

we do try to do is keep it to very controlled vocabularies that are widely used. Sir? Microphone. Turn your microphone on.

MEMBER TRANGLE: Microphone. My name --

(Laughter.)

MEMBER TRANGLE: My name is Mike, so I thought you were just talking to me.

(Laughter.)

MR. GOLDWATER: My apologies.

MEMBER TRANGLE: But is there anything to stop the Library of Congress subset that you're talking about from publishing different versions of the same thing so they compete and that the definitions are not the same?

MR. GOLDWATER: So --

MEMBER TRANGLE: Because if you're talking about CPT, for example, or just DSM-5 --

MR. GOLDWATER: Right.

MEMBER TRANGLE: -- and ICD-10, those are not aligned.

MR. GOLDWATER: I agree. So, the

answer is, yes, they can. That's sort of part of an issue. There's nothing on the VSAC right now that -- so, if you were to create a generalized anxiety disorder value set from DSM and there's already one that exists from ICD and SNOMED, the VSAC does not stop you from publishing that value set. You can go ahead and publish it.

MEMBER TRANGLE: So, we probably already have non-discrete value sets for depression, then?

MR. GOLDWATER: That's correct. You have many, many value sets --

MEMBER TRANGLE: And at --

MR. GOLDWATER: -- for depression.

MEMBER TRANGLE: And at some point, it would be really important to harmonize, get rid of duplication.

MR. GOLDWATER: So, we did take on that project last year to do that. And without actually getting into the nuts and bolts of that, it's very difficult to do, because there are really roughly over 55,000 value sets that are

currently published.

So, harmonizing what is already in there is difficult. So, what we are trying to do, then, is to create some sort of national way of looking at it, to make sure that they are national value sets that can be used. Yes?

MEMBER JENSEN: So, I work for the VA and we have had electronic health record for 20 years.

MR. GOLDWATER: All right.

MEMBER JENSEN: And we do our own performance measures. Do you think we use the VSAC language? Or what -- value sets? Do you know? I mean, I guess I should know the answer to this, not you, but just curious if you know that.

MR. GOLDWATER: I would imagine that you would, yes, unless you have some other way of clinically coding the information. But I know VistA does use the VSAC value sets. So, I don't know anyone that doesn't use those.

Sometimes they will create their own

and not publish them, and that's what we're trying to stop. Because, then, if they're not reusable, then we run into this problem where everybody keeps creating duplicative value sets. Yes, ma'am?

MEMBER MAZON JEFFERS: Maybe you're going to answer this later, but can you just talk a little bit about how this gets operationalized if you're using a value set that comes from a claims or billing information versus using a value set that comes from diagnostic information that might coming directly from an EHR?

MR. GOLDWATER: Right.

MEMBER MAZON JEFFERS: And how does that play out?

MR. GOLDWATER: So, it's a good question. And I think it's important to know that value sets are not -- I mean, they're primarily used in electronic measures, but they're used in chart-abstracted measures too. I mean, you have to represent the clinical content one way or the other.

So, in some cases, there's alignment between the value set that is used in the chart and the value set in the EHR, which is what we look for. If there isn't, then we have to try to map to what's been nationally used in the VSAC.

And the reason, again, why I want to reemphasize why we do this is, we're not -- we wouldn't stop a measure if it doesn't do it, but what we will do is, we will make the developer go back and use national value sets that are reusable.

It is -- in the committee that we had to try to harmonize this, that was one of their best strategies moving forward in trying to at least attempt to get to some common denominator of what value sets people will use.

Because if we continue on the path that we were on -- again, in depression, ironically, sitting here at the behavioral health CDP, was one that we focused on during our meeting and there are at least 300 variations of depression, just as value sets.

There's depression, there's general depression, there's major depressive disorder. I don't know what the difference is between those three, I'm sure there is, but if you're going to use it in a measure, you have to be specific about what clinical area you're focusing on and make sure that if another developer's going to do a measure like this, that they can use the same exact value set.

CO-CHAIR BRISS: So, I'd like to fairly quickly finish this --

MR. GOLDWATER: Right.

CO-CHAIR BRISS: -- and try to get on to the next measure.

MR. GOLDWATER: Sure.

CO-CHAIR BRISS: And so, Harold, quick comment, and then, let's see if we can finish this up.

MR. GOLDWATER: Okay.

CO-CHAIR PINCUS: Yes. I just wanted one very quick comment. Just, with regard to value sets and developments in sort of diagnostic

taxonomies, that actually the World Health Organization is in the final stages of developing ICD-11, which people may not realize that the rest of the world has been using ICD-10 for 25 years.

MR. GOLDWATER: Right.

CO-CHAIR PINCUS: And so, that -- and in that, that's being harmonized with SNOMED and being built off an informatics platform. Actually, I co-chair the Quality and Patient Safety Topic Advisory Group for WHO for ICD-11. And so, some of this will, over the next -- depending upon when the United States actually adopts ICD-11, is another issue. But --

(Laughter.)

CO-CHAIR PINCUS: -- that will -- but it's being done in a way that it's actually coordinated with the ICD-10-CM as well.

MR. GOLDWATER: Right.

CO-CHAIR PINCUS: But, anyway, I think we do need to move on to thinking about, so, how do -- what is the expectation for our review

today --

MR. GOLDWATER: Sure, right.

CO-CHAIR PINCUS: -- of this?

MR. GOLDWATER: So, and that sort of leads right in to what we're going to talk about. So, when you look at an eMeasure, it does have to be tested for reliability and validity. So, the requirement is that it has to be tested in at least more than one vendor or two, at least two.

So, developers need to test on the number of EHRs they feel is appropriate, it's highly desirable that you test in multiple systems. So, a question we often get asked is, well, I know an Epic System in Cleveland Clinic and I know an Epic System at, pick some hospital in New York, Columbia, Weill Cornell Medical College, and they both have Epic, is that one EHR or is that two?

That would be two. And the reason is is that every Epic installation is different. There are variations on a theme, there's a lot of similarities, but if you really study the

implementations of the system, there are differences. So, that would be considered to be more than one.

You have to indicate how the eMeasure specifications were used to obtain the data. Where did you get it from? What EHR? Did it align with the specs of the measure? And, again, you have to have it in the right format and that the data that is in that format can be implemented into that EHR system. Which, if you've done everything correctly, that shouldn't be a problem.

If the testing of an eMeasure occurs on a small number of sites, it could best be accomplished by focusing on patient-level data validity. The testing of level data elements require that all of them be tested.

At a minimum, the numerator, the denominator, and exclusions must be assessed and reported separately, which in the measure you're going to look at, it was. We do have some flexibility, understanding that not every

eMeasure is the same. We do have some standards we do have to adhere to, but we also do understand that, in certain cases, eMeasures are going to have something different to them.

So, we do take them on a case-by-case basis, I don't just look at them and say, well, this is good or this is bad, it's really, what are you trying to get at, can the measure be computed, does it produce the most reliable metric?

So, you should consult with NQF staff if you think you have another reasonable approach that will help you test for reliability and validity to ensure that your measure is at least getting a fair shake and representation at a CDP meeting. Okay? Any questions? Yes?

DR. LUSTIG: I think you --

MEMBER LARDIERI: There you go. So, it's -- I guess, for NQF in general, should NQF, I guess the question is, should NQF even be taking any measures in the future that are not eMeasures? I think if we restricted only

eMeasures, we'd have significant savings in administrative costs across the country.

And also, I think it would help to spur providing technology, appropriate technology, to those providers who don't have technology now, because if you're going to penalize or reward a provider, you need to make sure that they have the appropriate technology.

If this HQMF format is available, can be web-based, there's no reason why you should be doing anything paper anymore, ever, I think. So, I'm just wondering what NQF --

CO-CHAIR BRISS: So, thank you. I'd like to parking lot that, which is a general issue and not actually related to our committee work. So, if I can go on to David?

MEMBER PATING: So, I think the issue that affected our -- what will affect our rating is the Bonnie testing. So, I'm wondering if you can just explain that a little bit more? I saw many measures were Bonnie tested and it looked to me, there was either a thumbs up or a thumbs

down. Is there a score or a threshold value that signifies adequate Bonnie testing?

DR. LUSTIG: So, Jason, just to also be clear, the two eMeasures that this group will look at are both legacy eMeasures.

MR. GOLDWATER: So, Bonnie testing, the original design of Bonnie by MITRE was to test the meaningful use measures for feasibility or ensure that the measure logic calculated correctly, before they were actually tested within an actual EHR system.

So, the issue with legacy measures is that, some of the testing data that would have been available or the ability to test those measures, which were at one point chart-abstracted and then being moved into an EHR, was proving to be somewhat difficult.

So, the compromise was to be able to use Bonnie as a way to test the legacy measure to ensure the fact that the numerator and denominator could be populated appropriately, the exclusions were being able to be taken out of the

measure as needed, and that the metric calculated met the objective of the measure.

So, when we ask people to look at Bonnie, what we ask is -- you have to realize that Bonnie takes a synthetic test deck of patients. In other words, it is patients that you're creating, that you want the developer to create the most realistic set of patients. So, those individuals that would generally be in the population.

You do not want -- one of the first things is, did the developer just create a set of patients that the measure will obviously pass? If that's what they've done, and we usually catch that when they do, and that did not happen here, but when that happens, then we have to have them go back and actually create a more diverse and representative population, so that you can see the exclusions, the exemptions, the numerator and denominator being filled, and the calculation being met.

So, we then asked, take into account

the synthetic test deck of patients, does it represent the general population that would be put into this measure if needed? Is the measure calculating appropriately? Is it populating the numerator and denominator as it should?

Is the metric that it's producing mapping to the objective of the measure? And does that measure, overall, when produced, is that going to be an adequate representation of quality performance, based on the original design of the measure?

Because if it does, then the measure generally is meeting feasibility, because they will still have to have it in the appropriate format, they'll still have to map to a data model, it will still have value sets, all of those elements and pieces will still be in place.

But when we look at Bonnie, it is understandably, being a synthetic environment, is it as close to being representative of the population as it could be? In the particular case of these measures, they also did test them

in EHRs, it wasn't just testing in Bonnie. So, they did test in EHRs.

So, in addition to looking at the Bonnie testing, you can also look at the way that they were tested. And I think VistA was one of, if I'm remembering correctly, was one of the systems that was tested. So, you have the best of both worlds, you have it tested in Bonnie and you also have it tested in EHRs as well.

CO-CHAIR BRISS: Okay. So, one more comment, Harold.

CO-CHAIR PINCUS: Just related to the way in which we review these things today, your comment about the fact that it has to be done in two different EHRs, but that an Epic EHR in two different places counts, does that -- I'm concerned about that being sufficiently generalizable. And also, what if the two places are sort of, like, the sort of Jane Doe EHR versus Joe Schmo EHR that are rarely, if ever, used except in sort of idiosyncratic places?

MR. GOLDWATER: It's a good question.

The reason why we allow the different implementations of similar EHRs is because the evidence indicating the differences between implementations is pretty significant. And while it is highly desirable that they actually use two or more different vendors, sometimes the restrictions are going to be there that are not going to allow that.

So, if they are able to find two different EHR implementations that are the same vendor, they can show what those differences are and that they were adequately tested, then that does prove feasibility, if it matches all of the feasibility criteria.

Yes, in certain instances, it would be preferable to go to more vendors, and particularly in behavioral health, where there are smaller EHR vendors and not the Epics that you normally see in large hospitals. So, in that particular case, some of the testing does accompany that. In other cases, it does not.

But the requirement is, if they can

find two EHRs, get enough data to adequately test the measure, show that it was implemented and has produced the appropriate metric, then it has met the feasibility criteria.

CO-CHAIR BRISS: So, we still have on our plate to kind of do the eMeasure version of the last one that we just did. We're a little behind, I'd still like to see if I can move us through before the break.

So, does anybody have questions that are specifically related to the work we have immediately in front of us in terms of evaluating these eMeasures? And hearing none, let's try to fairly quickly go through this.

I'm hoping that this can go fairly quickly because many of the criteria are likely to be pretty similar to the non-eMeasure that we just went through. And maybe I'll start by trying to make -- does the developer, want to tee this up? Is there anything specific to the eMeasure version that you need to tee up that we didn't already talk about in the other version?

MS. SOMPLASKY: Not really, no. We're prepared to speak to the sites and vendors that were tested.

CO-CHAIR BRISS: So, I might like to propose, it appears to me plausible that we might be able to make the case that this is still important to measure and report, sort of regardless of whether it's an eMeasure version or another one, and I'd like to skip that, unless somebody has other ideas.

And I'd also like to skip that there's a likely measure gap, right? And so, does anybody want to object to that and talk further about --

DR. LUSTIG: So, just to clarify, we automatically assign the evidence rating that we voted on last time, because those are identical. There is a slight difference with gap, they do present data based on the EHR data, and so, we should technically have a vote on that.

CO-CHAIR BRISS: Okay.

DR. LUSTIG: But there doesn't need to

be discussion unless folks think there's a need for it.

CO-CHAIR BRISS: So, can you scroll down on the screen to the specific new data on the -- that came from these -- so, that's the data that we saw before? And this is the new EHR data from two places, right? Is that right?

DR. LUSTIG: Yes.

CO-CHAIR BRISS: So, that's the only new piece of information that we now have. So, would anybody like to further discuss that or can we move to a vote? I'm going to take silence as assent and let's move to a vote.

MS. QUINNONEZ: Voting is now open for the performance gap of Measure 3132: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

We are now voting on performance gap for Measure 3132. This is Preventative Care and Screen: Screening for Clinical Depression and Follow-Up Plan, the eMeasure. Option 1, high; option 2, moderate; option 3, low; and option 4,

insufficient.

All votes are in and voting is now closed. For the performance gap of Measure 3132: 57 percent voted high, 43 percent voted moderate, zero percent voted low, and zero percent voted for insufficient. Measure 3132 passes the criterion of performance gap.

CO-CHAIR BRISS: So, we do need to talk about reliability. Would one of the discussants like to volunteer? Thank you, Lisa.

MEMBER JENSEN: You're welcome. The group that looked at it determined that the Reliability does seem to be good. Some question about whether the varied EHR codes and ability to extract that information are the same.

Data elements clearly defined. Reliability was tested in two different practices, primary care and peds, and was determined to be good reliability. That's all I'll say.

CO-CHAIR BRISS: Anybody else have additions? Yes, Jeff?

MEMBER SUSMAN: I mean, just very briefly, if I'm looking at the right data here, they had four providers in MEDENT 22.0 and then a bunch in Centricity, and you just have to ask yourself, is that really strong enough data or support for reliability?

So, I mean, I would not say it's high by any means, just on the basis of that particular level of testing. I guess I would encourage CMS to be looking at trying to lure more durable testing and more durable sample of different EHRs when providing this to us. Because I look at it and say, well, okay.

CO-CHAIR BRISS: Anybody else want to speak to reliability? Yes, Shane?

MEMBER COLEMAN: I had a related question. I'm just curious if there was a rationale for why only those two EHRs and fairly small, I think, representative sample?

MS. SOMPLASKY: This Measure was initially supposed to be brought in front of NQF in 2018, so we were given a very short period of

time to get this ready for presentation today. It's voluntary for practices to participate in testing.

We reached out to more than 50 practices that represented ten different vendors and we actually had three sites immediately say yes. We had to discount one of the sites because they just weren't able to get the structured data that we needed. So, what you see are the two sites. And the Bonnie testing.

MEMBER COLEMAN: One other question is just, I'm not -- in some ways, to evaluate this, I kind of feel like I need some idea of what the market share of EHRs would be across the country, of sorts.

I mean, just because, like, looking at these, I have no idea if these two together represent one percent of the market share across the U.S., in which case, that would be more problematic then, I don't know, if it was 50 percent of EHRs, it would -- I would think it was more generalizable. Do we have any idea?

MEMBER LARDIERI: Centricity is probably about the third or fourth. You've got Epic, then you've got Cerner, then MEDITECH, and Centricity, and then a couple other ones fall in that. So, they're at the top ten at least, anyway.

CO-CHAIR BRISS: Raquel?

MEMBER MAZON JEFFERS: I just have a question about whether we're supposed to be thinking about reliability simply within the confines of the eMeasure or are we also supposed to be thinking about this in how much the Measure measures a similar thing in a similar way as the analog Measure, the non-eMeasure, so that there's consistency from the non-eMeasure to the eMeasure? Or are we just really thinking -- evaluating both those Measures to the question of, does it measure what we want it to measure?

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

CO-CHAIR PINCUS: I think it's also a question that would be fair from NQF, about

whether -- because that's actually an important issue.

DR. BURSTIN: It's actually a great question. And I think one of the reasons we wanted to make it clear that the eMeasure is actually in some ways a very distinct beast is, we've seen that rates of performance can be very, very different. Now, it may be that rates of performance are different because this is a better data source and it's closer to truth.

And so, I think we want to be able to not necessarily hold the eMeasure hostage to what the rates of performance may have been on a data source that may not have been as good, but at times, we're also seeing the opposite, which is that it is at times very difficult to get standardized data in eMeasure fields and so, the paper may actually have been closer to truth.

So, I think it's a great question, I don't think we have a clear answer that says, therefore, they have to be relatable. We're trying to put them side-by-side.

I think an assessment from the Committee about the quality of what you think you're getting in terms of the reliability of the Measure results on one versus another, as we think to the earlier comment about prioritization, would actually be very, very useful.

CO-CHAIR PINCUS: Yes. I mean, I think another reason might be that organizations and practices that have the capacity to report eMeasures may be higher performing. That may be another sort of issue with that.

But I think, I mean, so one recommendation to NQF is actually that -- if the intention is that these be reported as -- is the intention that these be reported, that is the non-eMeasure and the eMeasure, be reported combined or is two separate reports?

And so, it -- which kind of goes back to Mike's larger issue is that, are we now going to be having two different reports, one an eMeasure and the other a non-eMeasure, for every

Measure?

DR. BURSTIN: Again, we'd love to move away from having the older version of a Measure, if we think the new version of the eMeasure is truly best in class. I think we don't have confidence at this point that you could take the rates of performance on a Measure that's an eMeasure and the prior rates of performance on claims or paper and necessarily assume comparability.

So, I think we've indicated in some of our prior documents, you should not be comparing performance across different versions of Measures, since a lot of that could just be the measurement differences inherent to the data source.

CO-CHAIR PINCUS: So, Measures users should --

DR. BURSTIN: With caution.

CO-CHAIR PINCUS: Essentially, you'd be saying that, Measure users should be making a choice about whether to use an eMeasure or the

non-eMeasure?

DR. BURSTIN: Correct. And hopefully, moving to the eMeasure, if the eMeasure can meet your needs.

CO-CHAIR BRISS: Anybody else have comments about reliability? It looks to me like we can move to a vote, please.

MS. QUINNONEZ: Voting is now open for the reliability of Measure 3132: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient. For the reliability of Measure 3132: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

Okay, all votes are in and voting is now closed. For the reliability of Measure 3132: 39 percent voted high, 57 percent voted moderate, four percent voted low, and zero percent voted insufficient. For the reliability of Measure 3132, this passes this criteria.

CO-CHAIR BRISS: So, would one of the discussants like to talk about validity? Again, I'd like to focus on the pieces of the validity

testing that are new to this Measure that we haven't already discussed with the other version of the Measure.

MEMBER COLEMAN: I'm happy to comment on it. The Bonnie testing I think showed good validity, I think it was 100 percent in both. The face validity, I think was the same from last time, so, the 12 experts, I believe. Otherwise, a frequency question came up, but it came up last time. It looks like that's mostly it.

MEMBER PINDOLIA: This is more for education, for my part, but for the Bonnie test, is 22 denominator adequate? It just seems so low, but I don't know anything about it.

MS. JOHNSON: I'd love to hear what the developer has to say, but when they're doing testing using Bonnie, they want to make sure that their Measure works. So, they want to make sure that that logic works.

So, they have to have enough patients in their deck, in their simulated data set, to be able to hit all the different permutations, if

you will, of young people falling out, people with this exclusion or that exclusion falling out, and that sort of thing. So, as long as that has been done, I think it is an appropriate data set. And I would love to hear if the developers look at it that way.

MS. WELLS: That's correct. That's my understanding is, we want to have a sample of patients that represent the different iterations of possible performance, and that's kind of what we're getting at with Bonnie testing.

MEMBER SUSMAN: So, it is essentially looking at and reading the different pathways or conditions and verifying that you've been inclusive of those in setting up these synthetic patients or synthetic charts?

MS. WELLS: Yes, I think so.

MEMBER SUSMAN: Okay.

MS. SOMPLASKY: Gary, Gary Rezek, are you on the phone, because you created the test deck for this, do you want to speak to how you created it? Maybe Gary's not on the phone.

MR. REZEK: I created the original test deck. We've also, in the process of the meaningful use annual update, the test deck has been subject to external review by MITRE and Lantana, who have provided us feedback and we've responded to that.

There are some required criteria for the test deck, namely being that there is 100 percent coverage of all of the logic branches. There's also -- we try to meet other requirements to make sure we've tested, anywhere where age is used, that we've tested boundary cases, boundary cases with timing, test concepts with -- that have ranges or units, such as lab results.

That every criteria is tested to be met and unmet, so that if something's missing or something is contrary to meeting the numerator, that it's calculating the performance correctly. And so, there is a level of subjectivity, but there's also a pretty robust set of criteria that we try to meet in the review process.

CO-CHAIR BRISS: Harold, and then

Shane.

CO-CHAIR PINCUS: So, for the previous Measure, there was a validity testing of comparing the claims to what was in the charts with regard to follow-up. So, how would one do an -- it's unclear to me how -- the Bonnie testing isn't exactly the same thing.

What would be the equivalent of how to do that with this eMeasure, to look at whether -- did what was sort of in the EHR, which could have been a checkbox, whether the person actually did it, the provider actually did what they said they did?

MS. WELLS: So, we like to do data element validity testing when we can and I think that's what you're getting at, where we look at the EHR data and then we compare it to the chart with an actual manual review, so we might see things in notes that didn't make it into the Measure, that kind of thing. And, again, unfortunately, we came up against our timeline and weren't able to do that this time around.

CO-CHAIR PINCUS: I mean, one of the concerns that I have, while eMeasures are a great thing in terms of feasibility and efficiency, the risk is that they're checkboxes that are mindlessly checked. And so, I just would wonder about the ability to have some way of testing whether these events actually occurred in a meaningful way.

MS. WELLS: For the specific testing that we did, we go onsite and we do workflow analysis. And then, we go through each of the elements and we are looking in the EHR for how that is documented and for follow-up plan, asking them -- and I will readily admit that for the electronic Measures, follow-up plan is very difficult for vendors, whether you're talking Epic, who we've talked to many times about follow-up plans, or you're talking to GE.

MEDENT is actually one of the better ones to work with, because they've actually created structured fields in response to the testing we've done. But it is part of what we

are walking through when we are there onsite and looking at their cadre of patients that would have been tested.

CO-CHAIR BRISS: And at the median, if the worst thing we can say about this is that it's a checkbox Measure, even the checkbox isn't being checked a third of the time, right?

CO-CHAIR PINCUS: Yes.

DR. LUSTIG: And just as a reminder, in terms of NQF criteria, face validity and Bonnie testing are considered acceptable testing for validity.

CO-CHAIR BRISS: So, anybody else with things that haven't already been said? Let's try to vote on validity, please.

MS. QUINNONEZ: Voting is now open for the validity of Measure 3132: option 1, moderate; option 2, low; and option 3, insufficient. For the validity of Measure 3132: option 1 is moderate; option 2, low; and option 3, insufficient.

All votes are in and voting is now

closed. For the validity of Measure 3132: 78 percent voted moderate, 17 percent voted low, and four percent voted insufficient. This passes the validity criterion.

CO-CHAIR BRISS: So, would one of the discussants like to tee up feasibility for us?

MEMBER JENSEN: I can go. I think that we already kind of touched on that, in that the difficulty of documenting the follow-up is what the Committee identified as problematic.

CO-CHAIR BRISS: Anybody have other comments? Let's move to a vote.

MS. QUINNONEZ: One second.

CO-CHAIR BRISS: We were too efficient on that one for you, weren't we?

(Laughter.)

MS. QUINNONEZ: You were. All right. Voting is now open for Measure 3132, we're voting on the eMeasure approval for trial use for specifications. Option 1 is high, option 2 --

DR. LUSTIG: No, we're -- sorry, we're voting on feasibility.

CO-CHAIR BRISS: Feasibility.

MS. QUINNONEZ: We'll go back, sorry about that. There we are, now we're reading. Feasibility?

CO-CHAIR BRISS: Yes, feasibility. It's feasibility.

MS. QUINNONEZ: Voting is now open for the feasibility of Measure 3132: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient. For the feasibility of Measure 3132: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

All votes are in and voting is now closed. For the feasibility of Measure 3132: 35 percent voted high, 61 percent voted moderate, four percent voted low, and zero percent voted for insufficient. This passes the feasibility criterion.

CO-CHAIR BRISS: And usability and use looks pretty similar to the other version that we already voted. Would -- do any of the discussants feel like there are specific things

to this version of the Measure that we need to discuss under usability and use?

I'm getting heads shaking. Would anybody else on the Committee want to add anything about this version of the Measure? Let's move right to a vote on usability and use.

MS. QUINNONEZ: Okay. Voting is now open for the usability and use of Measure 3132: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient information. For the usability and use of Measure 3132: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient information.

All votes are in and voting is now closed. For the usability and use of Measure 3132: 30 percent voted high, 70 percent voted moderate, zero percent voted low, and zero percent voted for insufficient information. This passes the usability and use criteria.

CO-CHAIR BRISS: And it does not appear to me that we need to have a discussion on related and competing, so I think we can move

straight to the overall vote on this one.

MS. QUINNONEZ: Voting is now open for the overall suitability for eMeasure approval for trial use.

DR. LUSTIG: It's not a trial use.

MS. QUINNONEZ: For endorsement.

CO-CHAIR BRISS: Yes.

MS. QUINNONEZ: Let's redo this vote. We will now be voting for the overall suitability for endorsement, for approval endorsement of Measure 3132: option 1, yes; option 2, no. For the overall suitability for eMeasure approval for endorsement: option 1, yes; option 2, no.

All votes are in and voting is now closed. For the overall suitability for endorsement for eMeasure 3132: 100 percent voted yes.

CO-CHAIR BRISS: All right. I have failed in, my only role is to be a stern timekeeper and I have failed in that.

(Laughter.)

CO-CHAIR BRISS: Let's slightly shorten

the --

CO-CHAIR PINCUS: You had a handicap with the first one.

CO-CHAIR BRISS: We may be a little behind for our first one, we're not hugely behind. So, I'm going to give myself a partially successful. Let's shorten the break a little and we'll resume promptly at 20 after.

DR. LUSTIG: Thank you to our developers. Thank you.

(Whereupon, the above-entitled matter went off the record at 11:10 a.m. and resumed at 11:21 a.m.)

CO-CHAIR PINCUS: Okay. So, why don't we get started? And let's hear from the -- so, we're considering Measure 3207: Medication Reconciliation on Admission.

DR. LUSTIG: No, 3205.

CO-CHAIR PINCUS: Oh, excuse me, 3205: Medication Continuation Following Inpatient Psychiatric Discharge. Number 3205. And let's hear from the Measure developer.

DR. CAMPBELL: Thank you. Good morning. My name is Kyle Campbell. I'm Vice President of Pharmacy and Quality Measurement at Health Services Advisory Group. This is my colleague, Dr. Almut Winterstein, from the University of Florida, who has collaborated with us in the Measure development process for this Measure.

The Measures that we're going to be discussing today with you were developed for the CMS Inpatient Psychiatric Facility Reporting Program, which is a pay-for-reporting program that includes about 1700 inpatient psychiatric facilities nationally.

We developed the medication continuation Measure to specifically address gaps in treatment following inpatient psychiatric admissions for frequently occurring diagnoses, including schizophrenia, major depressive disorder and bipolar disorder.

For these disorders, medications are the primary mode of treatment. And furthermore

outcome data suggest that patients with these diagnoses who are not adherent to their prescribed medication regimens have a much greater risk of relapse and also the potential to either harm themselves or potentially someone else.

We found intervention studies that suggested that inpatient facilities could play an important role in the care coordination process for these patients and could impact med continuation rates.

In our Measure development process, we received input from a multi-disciplinary technical expert panel and we also conducted in-depth interviews with 20 diverse patients and caregivers who had had recent experiences in the inpatient psychiatric facility setting. Those individuals suggested to us that medication management was extremely important and that the Measure would help them in their decision making, where they sought care.

The Measure utilizes CMS

administrative data to calculate the percentage of patients with schizophrenia, MDD and bipolar who fill an evidence-based medication within two days prior to discharge or 30 days post-discharge.

It's important to note that, similar to other CMS Measures based on administrative data, the only patients that are included are those that are Medicare fee-for-service patients and in addition, those that have Part D coverage. For this particular population, about 75 percent, 74 to 75 percent of the patients have Medicare Part D, who are fee-for-service patients.

We did find a significant quality gap and variation in performance rates nationally, suggesting opportunity for improvement, and our data suggests that the Measure is highly reliable and valid and can be implemented with little to no burden for inpatient facilities. So, we thank you for your consideration of this Measure and look forward to the discussion.

CO-CHAIR PINCUS: Okay. So, the first

criterion that we're going to be looking at is evidence. So, does one of the lead reviewers want to take a stab at discussing the evidence issue? Tami?

MEMBER MARK: So, they develop a logic model, pointing out evidence from guidelines and structured reviews that support those guidelines that lack of adherence to medication leads to relapse and negative outcomes. And so, that supports the evidence that people who get discharged from a psychiatric hospitalization should get a follow-up medication.

CO-CHAIR PINCUS: Other comments about evidence? Anybody?

MEMBER PINDOLIA: So, one just comment, but I definitely agree that I think the Measure --

CO-CHAIR PINCUS: Little closer to the mic.

MEMBER PINDOLIA: Oh, sorry. I agree that there is other data to support, but the comment that I have is that, when I was trying to

look for studies to see if there is actual data to show the timeliness directly results in outcome differences, I couldn't find that.

However, from my own work with our systems, psychiatric institution and the discharge, I know that's an important part. So, I don't think that question could be answered as one thing that we were looking into, but I think the data -- I think everyone understands that.

CO-CHAIR PINCUS: Okay. Les?

MEMBER ZUN: So, I had a couple of questions, it's more a general understanding of the Measure. So, first of all, how are we measuring that they're taking the medicine? I mean, it's writing the prescription, it's that they picked up the prescription, right? And they picked it up either two days before or 30 days after the hospital stay, but how do we know that -- how is it connected to actually compliance?

DR. CAMPBELL: Yes, that's a really good question. So, it is a claim in Part D that indicates that the patient purchased or paid for

the medication, right?

And there's ample data to suggest that those adherence rates, meaning that if we monitor or track adherence in accordance with claims data and we look at things like medication possession ratios or we look at proportion of days covered, that those data are correlated directly to outcomes. And so, we believe the same data or evidence would apply here.

MEMBER ZUN: So, I understand that, but I guess the gap is, but we don't know they took it, they've just filled it. And --

DR. CAMPBELL: Yes, so those are correlated, but to actually get at the question of whether the patient took the medication or not, obviously would be burdensome from implementation from a facility perspective. So, this is our best proxy that the patient in fact took the medication.

MEMBER JENSEN: We also don't know if they took it correctly. I mean, they may take it, but then they don't take the right dose --

MEMBER ZUN: They gave it to the dog.

MEMBER JENSEN: -- they decided, oh, it's too expensive, I'm just going to take half a dose.

CO-CHAIR PINCUS: Rhonda?

MEMBER ROBINSON BEALE: I think you're all making very good points. But one of the things that's unique with behavioral health, and particularly with a severe mental illness, is that you have injectables. So, I'm wondering whether or not there's a separate Measure for injectables versus oral administered drugs.

DR. CAMPBELL: So, thank you for that question. And, yes, we did consider long-acting injectables as part of the Measure, so they qualify for the numerator component.

And since most long-acting injectables have a duration of action less than a month, it's presumed that, even if that was administered during the inpatient stay, that within the 30-day follow-up period, that there would need to be another administration.

Which is -- and that's captured, actually, in the Measure, in Part D and Part B, because some of the injections are given in the physician office and billed to Part B, so that's part of the Measure numerator.

CO-CHAIR PINCUS: Mike?

MEMBER TRANGLE: My understanding is that this is limited to people with Medicare who have Part D, is that correct?

DR. CAMPBELL: That's correct.

MEMBER TRANGLE: Which, I mean, it seems like an important Measure, it's necessary but not sufficient. But in many ways, I'm also concerned about people that have other insurance types and at least several of our hospitals, we find that there are a lot of people being discharged on MA pending. They don't have medical assistance yet, but they've applied for it, or no insurance.

And we also find that we have a lot of patients that don't get their meds because of prior authorization kinds of things, they're on a

certain formulary inpatient and the med has to be changed to a different alternative, comparable kind of thing.

And in some sense, I don't have a good sense of if your sample is representative of the real world and the complexities and the things that get in the way. And if you could comment on that?

DR. CAMPBELL: Sure. So, when we evaluated the data, obviously being a Medicare administrative data source, we're limited to those patients with fee-for-service coverage. But in terms of the prescription drug coverage, about 74 percent of our patients in the cohort have prescription drug coverage under Part D.

And Part D is actually really pretty restrictive with plans in terms of the formularies related to behavioral health and that really substantially, in their language, in the Part D Call Letter, substantially all medications in the behavioral health class have to be covered.

So, basically, we're looking at patients in this Measure that have coverage and if they were low-income patients, they would also be eligible for a low-income subsidy. And, in fact, when you look at the follow-up rates in this Measure, the follow-up rates are, med continuation rates, are actually higher in the dual-eligible population.

And we hypothesize that's because they have higher financial assistance related to the medications being filled. So, we don't really have anyone in this Measure at this point that doesn't have coverage or would fall into those gaps and that sort of thing.

CO-CHAIR PINCUS: Any other questions?

MEMBER SPERLING: Very briefly.

CO-CHAIR PINCUS: Yes?

MEMBER SPERLING: So, CMS asked for this, correct? Or CMS was involved in development of this?

DR. CAMPBELL: Yes, correct. We're contracted by CMS.

MEMBER SPERLING: And the question is, are they going to harmonize this or use this for their Star Rating Systems for the Medicare Advantage and Prescription Drug Plans? Do you know?

DR. CAMPBELL: At this time, I'm not aware. There's no plans or there haven't been any discussions with the Star Rating program.

MEMBER SPERLING: Okay. Because the reason I mention it, that's a very powerful tool that CMS has, because if plans fall below the Star Rating for a certain period of time, they're actually booted out of the program. It's something that really matters to the Prescription Drug and Medicare Advantage Plans in Part D.

DR. CAMPBELL: Okay, thank you.

CO-CHAIR PINCUS: Tami?

MEMBER MARK: I just want to clarify, the inclusion criteria is not only that they have Part D, but they have to be enrolled in Part A, B, and D at least 30 days post-discharge, is that correct?

DR. CAMPBELL: That's correct.

MEMBER MARK: So, all of Part C is out?

DR. CAMPBELL: Correct.

MEMBER MARK: And the duals, would you -- if the duals have a hospitalization, are they going to be picked up? Because that'll be under Medicaid?

DR. CAMPBELL: The duals, yes, the duals are included in the data set, yes.

MEMBER MARK: But is their inpatient admission going to be picked up, because that part would be picked up under Medicaid, not Medicare?

DR. CAMPBELL: It's picked up on Medicare, I believe, because the Medicare is the primary payer.

MEMBER MARK: Medicare is the primary payer?

MEMBER PINDOLIA: And the Medicaid-only population?

DR. CAMPBELL: We don't have the Medicaid-only population in the Measure.

CO-CHAIR PINCUS: So, this is essentially a Medicare Measure?

DR. CAMPBELL: Correct.

CO-CHAIR PINCUS: With --

MEMBER TRANGLE: I think the obvious question that wasn't asked was, was there thought given to including MA or other ones, to make it a broader Measure and have more impact on overall patient flow through hospitals?

DR. CAMPBELL: Yes, I think that's a good question. Unfortunately, to limit the Measure to the administrative data, we don't have access to the administrative data for Part C, so we don't have the ability to do that. So, that was the limitation, yes, in the development.

CO-CHAIR PINCUS: Any other comments, questions with regard to evidence? If not, we're ready to vote.

MS. QUINNONEZ: Voting is now open for the evidence of Measure 3205: Medication Continuation Following Inpatient Psychiatric Discharge: option 1, moderate; option 2, low; and

option 3, insufficient.

DR. LUSTIG: And just to clarify, high is not an option here, because in the submission, there wasn't a clear review of the quality, quantity and consistency of the evidence.

MS. QUINNONEZ: All votes are in and voting is now closed. For the evidence of Measure 3205: 91 percent voted moderate, nine percent voted low, and zero percent voted insufficient. So, 21 out of 23 voted for moderate, two people voted for low, and zero voted for insufficient. This passes the evidence criterion.

CO-CHAIR BRISS: Do we -- it's interesting, do we have an opportunity to send messages back to CMS? I thought I heard around the table that there was a fair amount of support for seeing if the denominator of this Measure could be broadened to a broader population. Is there -- did I hear enough of us saying something like that, and if I did hear something like that, do we have ways to send that back to CMS?

DR. LUSTIG: Yes. And also, when we get to our criterion that have to do with specifications, we can do that.

DR. BURSTIN: And it's broader than mental health, obviously, this whole issues of measures for fee-for-service versus measures for MA is a problem and that the limited data source for a developer like Kyle is just -- we just don't have it. But, ideally, you would expect those to be the same measures done in the same way.

CO-CHAIR PINCUS: Just a question, are there any -- is there precedent for any other, in other areas of medicine, where there's a measure for medication follow-up post-hospitalization?

DR. CAMPBELL: I don't believe so, I think this is the first time for follow-up for medication continuation.

CO-CHAIR PINCUS: So, for example, for follow-up in terms of getting an additional prescription for a beta blocker after an MI or something, there's nothing like that?

DR. CAMPBELL: A beta blocker after MI, there is, but that measure was retired after it exceeded the --

CO-CHAIR PINCUS: Yes, but I thought that was at discharge and not for the 30 days after.

DR. BURSTIN: It was a HEDIS measure that looked at six-month use --

CO-CHAIR PINCUS: Oh, it was six-month use?

DR. BURSTIN: -- as well. But it's also been retired for --

DR. CAMPBELL: Right. And there are several that are -- look at long-term adherence to medications, over a year period of time.

CO-CHAIR PINCUS: Okay. So, let's look at the next issue of gap. Do one of the people on the reviewer list want to comment on the gap issue? Tami?

MEMBER MARK: I will, unless someone else wants to. I think the gap relates to the performance, is that right? And so --

CO-CHAIR PINCUS: Yes, performance gap.

MEMBER MARK: -- the distribution of the scores, the median, meaning the percent that got the medication post-discharge, was 80 percent. The bottom tenth percentile was 67 percent. The top ninetieth percentile was 88 percent.

CO-CHAIR PINCUS: Yes. I just, let me take step away from being Chair to say, I was actually surprised at how high it was.

MEMBER MARK: I was too.

CO-CHAIR PINCUS: Yes. That -- because this measure, obviously, would be highly related to measures of follow-up after hospitalization, which have a much lower rate. And also, other measures of treatment adherence over the course of a longer period of time have a much lower rate.

And so, I was just wondering whether, in the testing, there was something that might have skewed it, either because of the selection of people that maybe have more access.

DR. CAMPBELL: Yes, I think that's a good hypothesis. That, basically, we carefully crafted the eligible population and we also paid close attention to the exclusions. So, we're excluding patients here that might be relatively or absolutely contraindicated to a given medication.

So, I think, we have just patients in this population who probably don't have any access issues. They have full prescription drug coverage. If they're low-income, they also have assistance. So, I think that's why you're seeing higher rates than what you might see in the general population.

MEMBER MARK: But that's the population that's included in the measure. You're just saying that you limited it to the people included in the measure.

DR. CAMPBELL: I'm sorry?

MEMBER MARK: If I understand your point, you tested it based on the criteria in the measure, so you're not saying that when the

measure gets implemented, the rate is going to go down?

DR. CAMPBELL: Right.

MEMBER MARK: This is -- because you actually tested it on over 1,000 inpatient facilities, it was a very large sample. So, I don't think it's a weird, small sample.

CO-CHAIR PINCUS: Right. No, but it's the way the denominator is defined.

MEMBER MARK: No --

CO-CHAIR PINCUS: The fact that you --

MEMBER MARK: Well, it's still many, many, many -- most people with Part B coverage who get admitted to an inpatient psychiatric hospital are going to show this score. This is not --

DR. CAMPBELL: Yes.

MEMBER MARK: -- an unusual score. When this gets put out into the real world, we're going to expect the same scores.

DR. CAMPBELL: Right. No, yes, I think so, but it's not everybody who gets discharged

from a psychiatric hospital.

MEMBER MARK: It is pretty much -- they tested it on over 1,000 --

CO-CHAIR PINCUS: No, but it's only people with Medicare.

MEMBER MARK: Right. This measure is only -- right.

CO-CHAIR PINCUS: Yes.

MEMBER MARK: That point is correct.

CO-CHAIR PINCUS: Yes, that's what -- that may be --

MEMBER MARK: But that's not this measure, this measure is only people in Medicare.

CO-CHAIR PINCUS: Yes.

MEMBER MARK: I just -- I think the point is, people have pretty good continuity of care and access to medications. I mean, if you look at the percent of the population taking psychiatric medications, it's 15-20 percent, so it actually doesn't really surprise me that it's pretty high.

When you look at the percent of people

getting therapy or treatment outside of medications, it's pretty low. So, to me, this data is pretty consistent with all the other data we see around psychiatric, maybe getting a little away from our main point.

CO-CHAIR PINCUS: Okay. Rhonda? And then, let's --

MEMBER ROBINSON BEALE: Yes. I was just going to comment on your statement that this is surprising since the post-discharge follow-up measure is always so low. The problem with that measure, very familiar with that, Harold, is that the measure only measures follow-up with a behavioral health provider.

And most geriatric patients go back to their primary care physician or their medical physician and that's generally the one who is managing them. So, I would expect there to be a great disparity between the two measures.

CO-CHAIR PINCUS: Vanita?

MEMBER PINDOLIA: So, I understand that the range in the gap is only 66 to 88 and it's on

the high end, and that's good. But the concern I have is, did you look further into see -- and divide the population of -- if the gap was different if they were discharged to a group home versus individual home? I understand homeless shelter probably is not a population with the Part D, but we definitely see a difference when they go to group home versus individual homes.

DR. CAMPBELL: Yes. So, that's also a good question, and I saw that in the prior comments, so I had somebody run it before we came to the meeting. But not specifically to address your question.

So, in the measure, we only include those patients that are discharged to home or home health. All the other discharge disposition codes in the claims, of which there are many, hospice, SNF, all those types of things, they're excluded.

And I did look at a comparison between home and home health, and home health has higher rates of medication continuation, as you would

expect. But those discharged to home would be more reflective of the mean in the measure.

CO-CHAIR PINCUS: Mike?

MEMBER TRANGLE: We've had sort of a smoldering, I wouldn't say raging, controversy in our system about whether many of these folks, we should fill their scripts with the hospital outpatient pharmacy, so they leave with the pills, versus trusting them and their families to get them in their local pharmacies. Did your data answer that question? I didn't -- I wasn't part of the Committee review of this, so you wouldn't have pre-seen this.

DR. CAMPBELL: That's also another great question. So, one of the reasons, when we originally specified the measure, we were looking at the time frame for the follow-up to be post-discharge, right? So, 30 --

MEMBER TRANGLE: But we don't write them before two days before discharge.

DR. CAMPBELL: Yes, yes, but what we identified in the testing in some of the

facilities that we worked with were kind of those innovative programs where there's medications at the bedside, where they're delivered basically to the facility.

And there was some concern that, if we didn't back that date up a little bit, we -- those prescriptions might have been dispensed by the pharmacy a day before discharge or two days before discharge.

So, that's why we changed the window, the follow-up window, to allow that. And just to note, so, patients that have Part D, that's an ambulatory benefit, so it has to come from the ambulatory pharmacy.

MEMBER TRANGLE: The hospital outpatient pharmacy?

DR. CAMPBELL: Right, correct.

MEMBER TRANGLE: That they get right before discharge and they leave with?

DR. CAMPBELL: Yes.

MEMBER TRANGLE: But you don't -- you didn't look at that to see if that was a bit more

successful approach?

DR. CAMPBELL: We didn't look at it in terms of any data, we just had anecdotal reports that it was occurring, that there was filling prior to discharge. And so, that's why we changed that window.

CO-CHAIR PINCUS: Yes. And they wouldn't be able to tell whether it was more effective, because that's included in the measure. Other comments with regard to performance gap?

MEMBER MARK: Well, I just, I had a question about the testing of that. Because under the data element validity testing, you say, few discharges included provision of medication at discharge. So, that sounds like you did do some testing, although, I was wondering what the percentage was.

DR. CAMPBELL: So, what we meant by the provision of medications at discharge is, we had our physician reviewers look to see if there were any samples or anything that wouldn't be captured

in claims that were dispensed by the facility. So, like a one- or two- or three-day supply. But that wasn't the case, we didn't identify that.

CO-CHAIR PINCUS: What about, actually related to that, what about patients who have had longstanding conditions that may have a reserve of medication at home, when they return home?

DR. CAMPBELL: That's a good question, too. So, we did look at a distribution of the day supply on the prior fills, prior to the inpatient hospitalization. And, predominately, they were a 30-day supply. So, we believe that allowing a 30-day window post-discharge would require that the patient --

CO-CHAIR PINCUS: Right. Although, if they were hospitalized, it might have been because they weren't taking that medication before.

DR. CAMPBELL: True, yes. So, we think, but we think the 30-day window is probably the best approach, given that they may have some stockpiled supply, which means they may not fill

it immediately post-discharge.

CO-CHAIR PINCUS: But it does sort of -- if you consider that and also consider that some proportion of people may have problems with the medication once they leave the hospital and may, in fact, come back to a provider, who might change or eliminate the medication, that would make the gap less --

DR. CAMPBELL: Right.

CO-CHAIR PINCUS: -- than what you found. So, there might be somewhat less of a gap than it might appear. Other items about the gap?

MEMBER JENSEN: I just had a question. Are you looking at the clinical setting that the person is receiving their prescription from? Is it in the primary -- I think this is to Rhonda's question, is it -- are they receiving it from a primary care physician or a specialty care physician?

DR. CAMPBELL: So, the idea here is that these would be prescriptions that were prescribed at discharge from the inpatient

psychiatric facility. Right, and it's whether they filled it within 30 days post-discharge.

CO-CHAIR PINCUS: Well, what if they went the next day to see a provider and that person did, it would also be captured there too, right?

DR. CAMPBELL: Right. Yes. It doesn't have to be from a fill from the inpatient facility. If they saw a provider and that provider wrote a prescription and they filled an evidence-based medicine within that drug class, then the facility would get credit for it.

CO-CHAIR PINCUS: So, just, validity, having to do with sort of the accountability and the -- of the -- and this may be something that comes up later in terms of usability, but in terms of the control that the hospital has, from the perspective of the developer, you're saying that the hospital could basically have control of this measure by virtue of providing the prescription at the point of discharge?

DR. CAMPBELL: Yes. And other

practices, other interventions, like appropriate discharge planning and patient/caregiver relationship to establish the importance of the use of the medication. So, we did find evidence that there's several interventions that can increase the adherence that the patient would take the medication post-discharge.

CO-CHAIR PINCUS: Right. Although, if they solve the problem by giving the prescription, that could potentially undermine the connection to an outpatient provider.

DR. CAMPBELL: Yes, but --

CO-CHAIR PINCUS: In terms of, the patient would not necessarily have a need to go see the outpatient provider, if they already have the medication.

DR. CAMPBELL: I mean, I think that would be part of the discharge planning, right, and process from the inpatient facility. The recommendation is, yes, here's the prescription you should be taking, but you need to follow-up with your PCP or your psychiatrist.

CO-CHAIR PINCUS: Right. Because of the motivation of the individual, oh, I'm running out of my prescription, I got to go see somebody.

DR. CAMPBELL: Yes.

CO-CHAIR PINCUS: Which might also account for the differences between the two measures, of the follow-up after hospitalization measure. Any other comments about the gap? Okay. We're ready to vote.

MS. QUINNONEZ: Voting is now open for the performance gap of Measure 3205: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient. For the performance gap of Measure 3205: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

All votes are in and voting is now closed. For the performance gap of Measure 3205: 30 percent voted high, which is seven people, 70 percent voted moderate, which is 16 people, and zero percent, zero individuals voted low, and zero individuals voted insufficient. This

measure passes the criteria for performance gap.

CO-CHAIR PINCUS: Okay. Let's move on to the next item, which is reliability. So, one of the lead reviewers want to comment on that?

MEMBER JENSEN: The reliability, the group looked at, was that it was acceptable, adequate. I don't think there were any particular questions about it.

CO-CHAIR PINCUS: Any other discussion with regard to reliability?

MEMBER MARK: I'll just point out that they used the signal-to-noise ratio and if I understand this right, they found that you have to have at least 75 discharges in order to be able to participate, to get a reliable measure.

DR. CAMPBELL: Yes, that's correct.

CO-CHAIR PINCUS: And that's -- and most of these facilities have that.

DR. CAMPBELL: Yes, about 1200 out of about 1700 had at least 75 discharges. And we did use a two-year measurement period to increase the number of facilities that could report, or

could be reported on.

CO-CHAIR PINCUS: Vanita?

MEMBER PINDOLIA: So, my question was more based on the type of drugs that we're looking at for the numerator to qualify. So, like the SSRIs, which are heavily, all generic and part of many $4 program, did you look at how many patients are getting their prescriptions that aren't using their prescription card and the claims processor because of that?

And that can vary from region to region. Just in Henry Ford Health System, I did that analysis and five years ago, it was already at 12 percent would not show and it's been growing steadily since then. And especially as now Meijer is also supplying it, Walmart, Costco, there's multiple.

DR. CAMPBELL: Yes. I think that's a good question. And we have developed other measures of adherence, NQF-endorsed measures of adherence for behavioral health.

My understanding, though, is under

Part D, that Medicare has been actively working with the folks that administer these programs and that there's some, I guess, capture, additional capture of the data. But I don't know to what extent that's occurring at this point.

MEMBER PINDOLIA: I helped lead for the Part D metrics at the plan in Henry Ford Health. And there's been no boost effect, there's been nothing, even if the data's been submitted, that 12 percent of our population did not show their cards. And now, it's growing to 18 percent.

So, I have not seen that and I'm just concerned as we keep having more and more drug adherence-related measures and we're continuously having more drugs that go into that program, so I just want to -- as we add more, I think there should be a further discussion or dissection of the data to understand that.

DR. CAMPBELL: Yes, I think that's a really good suggestion and something that we'd definitely take back to CMS.

CO-CHAIR PINCUS: Any other comments

about reliability? Okay. I guess we're ready to vote on reliability.

MS. QUINNONEZ: Voting is now open for the reliability of Measure 3205: option 1, high; option 2, moderate; option 3, low; option 4, insufficient. For the reliability of Measure 3205: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

All votes are in and voting is now closed. For the reliability of Measure 3205: 26 percent voted high, six individuals, 74 percent voted moderate, 17 individuals, zero percent voted low, and zero percent for insufficient. This measure passes the reliability criterion.

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

MEMBER JENSEN: So, with validity, I think we're back to that question of, patients may be picking up their prescriptions, but are they in fact taking them, and are they taking them appropriately as prescribed?

CO-CHAIR PINCUS: Other comments with

regards to validity?

MEMBER PARISH: I had one question in terms of, if somebody didn't really get their meds changed, that they're on a mail-in order plan that they already have four months of med at home, they wouldn't be picking up another med.

DR. CAMPBELL: Yes. So, we did an analysis of the frequency of days' supply for the patients in the cohort and the overwhelming majority of patients' prior fill to an inpatient hospitalization was 30 days.

CO-CHAIR PINCUS: And even if they were on a regular supply, that would have popped up anyway. But what if their medication was changed?

DR. CAMPBELL: If their medication was changed, then obviously they would have to fill --

CO-CHAIR PINCUS: Right.

DR. CAMPBELL: -- whatever was prescribed at discharge, which means that they would need to have a fill within the 30 days.

CO-CHAIR PINCUS: Right, but what if they kept on taking the old one or they got -- yes, I'm just sort of -- that's sort of -- people do automatic refills. If they got changed to a different medication, but they're getting the continuing refill is from the old medication.

DR. CAMPBELL: Oh, I see. So, if under their mail-order system, they went ahead and refilled from their prior medication, yes, I don't think that there would be any way to address that or capture that within the data, because we do allow any evidence-based medication within the specific drug classes related to the indications.

CO-CHAIR PINCUS: David?

MEMBER PATING: This is a really basic question. So, in terms of this validity question, is the validity question, did they get the prescription filled or are they taking the medicine, which was related to the gap?

So, it wasn't quite clear to me. Are we -- filling and taking the medicine is not the

same and I just wanted to make sure I understand the actual validity question that we're trying to resolve.

CO-CHAIR PINCUS: So, the validity goes on, obviously, at several different levels. One is the validity of the measure itself in actually measuring what it purports to measure.

And then there's the validity of the concept as it's being applied, in terms of, really, are people actually ingesting the medication, because that's where the evidence is that suggests the measure actually results in better outcomes?

MEMBER PATING: So, could you speak to that second linkage issue? Did you have data on that?

DR. CAMPBELL: So most of the data, because most studies use a proxy for adherence, which is the filling of the prescription, which is what we use, so most of the outcomes data related to whether the patient had an adverse outcome, if they weren't filling the medication,

is not related to a patient-reported measure, let's say, that, I took the pill every day for 30 days.

It's actually claims data that's utilized to figure out if they actually filled the medication. So most of the data, that's the way most of the evidence is characterized.

CO-CHAIR BRISS: So just on that point, it seems to me that filling a prescription sort of may or may not relate to taking the prescription, but if you're not filling the prescription, then you know that the person is not taking it, right? And so --

DR. CAMPBELL: Yes.

CO-CHAIR BRISS: -- there's a fair amount of gap in even filling the prescription. So you always have this tension about what can be measured and how good of a proxy it is. But I think this does tell us something important about continuity of care.

CO-CHAIR PINCUS: At some point, there may be an eMeasure that actually is using

wearables to determine, but we're not there yet. Rhonda? And then Vanita and Tami.

MEMBER ROBINSON BEALE: I may have jumped too far ahead, I was going to threats to validity, would that be --

CO-CHAIR PINCUS: Oh, okay. Vanita?

MEMBER PINDOLIA: I had a question on validity test reporting. So, I see the data element validity testing, and I understood that makes sense; you're trying to validate what this measure was doing.

But the measure score validity test, I just had concerns on that, because when that correlation was done, there was a large effect for the 30-day follow-up, but definitely not for the seven-day or the all-cause readmission.

What is -- if your measure is just, they filled a script or not, what is the jump to saying, did this decrease all-cause readmission?

I'm trying to understand the need for that validity test and where does that -- is this going to be used as measure, eventually. I'm

always concerned -- it's going to eventually become a 5-Star, and is it then going to become 3 point, because it's an outcome measure all of a sudden, instead of a process measure, which would be a 1. And so that's why my question.

DR. CAMPBELL: Okay, thank you for the question. So we actually developed and the measure received endorsement, the IPF All-Cause Readmission Measure, and our hypothesis was that if patients didn't fill their medication immediately post-discharge, that they'd be more likely to have a readmission within the 30-day window.

And, in fact, that's what we found, when we looked at the Spearman's rank correlation. I mean, granted, the correlation isn't large, but it's in the direction that we expect and supports the hypothesis.

And then if you look at the follow-up after hospitalization, the correlation gets stronger between seven-day and 30-day, which in our mind, if we're trying to show that this is a

measure of quality of care, it does correlate well to the existing endorsed measures related to the same construct, which is coordination of care.

And it's really, actually, exciting to be able to do this, because a lot of times we don't have the data, or we don't have the right measure to compare it to, but in this case, we feel that these measures do, in fact -- if you follow up with your outpatient provider, you're more likely to fill a prescription.

MEMBER PINDOLIA: So it's looking at correlation to coordination of care?

DR. CAMPBELL: Yes.

MEMBER PINDOLIA: Okay.

CO-CHAIR PINCUS: Okay. David?

MEMBER EINZIG: So, just for clarification, when we vote on validity, are we, I understand what we're measuring, but are we voting on the validity that simply the medications are getting filled, or the spirit of the measure of: are the patients taking the

medication?

CO-CHAIR PINCUS: Howard, you want to -- yes. I mean, my sense is that it's both.

DR. LUSTIG: It's both, yes.

CO-CHAIR PINCUS: Yes, it's both. Yes, it's both, obviously, but in many cases, we're looking at data that are sort of surrogates to actual endpoints, that are intermediate endpoints. That's the reality in terms of the feasibility of this.

CO-CHAIR BRISS: So David, I would say, so the way I would frame that up is that it's both -- are you actually able to measure what it purports to measure? So is it validly measuring whether people filled scripts on the one hand? And then is that measure an acceptable proxy for actually taking the med? It's both.

CO-CHAIR PINCUS: Okay. Raquel?

MEMBER MAZON JEFFERS: And then, I just would go one step further, and then, is that measure a measure of continuity of care? Because that's where I'm falling off the edge, because I

don't see it as a continuity measure. Because you are saying the majority of the prescriptions that are being filled were written by the inpatient hospital, not by a community provider.

DR. CAMPBELL: Right. So in order to ensure, particularly in this population, that patients fill their medications post-discharge, it's sort of that first step in continuity. And so I think -- and it's also the step that we feel like, since the IPFs most likely wrote the prescription, that we can hold them accountable for it, right?

We're not asking them to stay accountable for what happens six months from now. We're asking them, for the prescription you wrote, did your patient go and fill that medication?

CO-CHAIR PINCUS: It's not a measure of coordination of care. Yes, Rhonda?

MEMBER ROBINSON BEALE: I'm saying, it's more a measure of transference, if the patient had a positive transference to the

treatment and therefore wanted to fill their prescription because they understand the importance of it. It's kind of like that.

CO-CHAIR PINCUS: Or whether the hospital just gave them the medication as they were going out the door. I mean, that's -- it may not be the patient went out and filled it in the case of this medication.

One question I had related to validity is the issue of sociodemographic risk adjustment. I saw that African American patients performed worse on this, and whether there's a need -- and I can imagine that safety-net hospitals may have more of a problem in, for -- those patients that are being discharged to other facilities, may have more of a problem in getting follow-up for those patients and patients having sort of the organizational capacity and other sort of socioeconomic supports to go get their prescriptions. How might that be impacted?

DR. CAMPBELL: So it's generally a process measure. So for NQF, we generally don't

risk-adjust process measures.

CO-CHAIR PINCUS: Well, I thought that there's a period of time in which we're looking at that.

DR. BURSTIN: Right. So during our trial period, we are looking at all measures to see if there's a conceptual basis for adjustment and thinking about whether that makes sense in this particular measure. We don't usually look at process measures, I think, as Kyle correctly points out.

I will say, though, that adherence is sort of a funny process/outcome, it's not clearly a process. I think many would think of it probably as an intermediate outcome, which certainly would be eligible for this.

I will also say that the Pharmacy Quality Alliance has a series of adherence measures that they are actively in the process of adjusting for SES, so there is some work already happening here.

CO-CHAIR PINCUS: So if we endorse

this, does that mean we endorse it even though there's not a specific sociodemographic adjustment built into it, or are we encouraging that one be built in?

DR. BURSTIN: I think it would be reasonable to hear from the developers to see if they've thought about it. Do they believe there's a conceptual basis and whether there could even be data to look at whether it's possible? It's difficult to -- I mean, again, there's no current risk adjustment for the measure.

CO-CHAIR PINCUS: Right.

DR. BURSTIN: So it's a little hard to conceive of throwing in additional variables for a risk model that does not yet exist. That being said, you may suggest this measure be stratified for the populations most at risk to ensure we're understanding what those differences are.

DR. CAMPBELL: Yes. In general -- and we participated in the sociodemographic pilot related to the IPF All-Cause Readmission Measure.

In general, we don't look to risk-adjust for race/ethnicity, because we don't want to risk-adjust away any potential disparities in care.

CO-CHAIR PINCUS: Right, but what about sociodemographics more broadly?

DR. CAMPBELL: Sociodemographics, we actually didn't find in the IPF All-Cause Readmission Measure that they were more explanatory than the clinical comorbidities that we used, so we didn't recommend an SDS adjustment.

CO-CHAIR PINCUS: Right, but we're talking about --

DR. CAMPBELL: This measure, yes, I'm getting to that. So for this measure, as I mentioned, because it's a population of patients that have prescription drug coverage, and because those patients also would be eligible for low-income subsidy, I don't think there's really any barriers to access there that would necessitate us doing any types of adjustment.

And in fact, when we looked at the

data, the dual-eligible population, which is generally a good marker in claims of lower socioeconomic status, they have higher rates of med continuation, rather than lower. So I don't think there's anything other than the disparity we noted for race/ethnicity where we might consider stratification.

CO-CHAIR PINCUS: Rhonda?

MEMBER ROBINSON BEALE: So with that being said, I have a question about the exclusion of patients with delirium secondary diagnosis and dementia secondary diagnosis, because I'm not quite clear I understand why.

DR. CAMPBELL: So one of the things we did as part of our exclusion analysis was to look and say, is there any reason why a patient would not have an indication for a given medication in the class?

And for patients with schizophrenia and dementia, there is a black box warning, as many of you are probably aware of, and I think the APA has come out with a recent statement

about how to handle that, and it's a patient-by-patient sort of risk/benefit tradeoff.

And for that reason, our technical expert panel said we should probably just take those out because we can't ascertain that risk/benefit in claims. And delirium is a similar thing; some of the medications in the measure are contraindicated for delirium.

CO-CHAIR PINCUS: Okay. Vanita?

MEMBER PINDOLIA: I'd like to further challenge the sociodemographic question that was raised by Harold, because when you compare the dual-eligibles, they have pretty much co-pay. For Part D, when they're in the donut hole, and if they do need one of the second or third generation atypicals, that 40 percent of the brand that they have to pay is almost impossible for many, and they don't qualify for the low-income subsidy.

So I know from our experience with COPD, that's a huge problem we have, because of the cost of inhalers, and they're not generic at

all; there's no generic available. And if they have a 20 percent on their nebs, we cannot do an alternative, other than give it away free. So I don't think you can extrapolate your dual-eligible population data to this because this is more pure Part D where they won't have that.

DR. CAMPBELL: Do you -- I mean, I think if we were to explore this in the future, do you have a variable that would be available in claims data that would allow us to do that that is not the dual-eligible variable for Part D? I think that really is -- that's really the question is, in the claims data, what data do we have as a proxy for sociodemographic status?

MEMBER PINDOLIA: Well, we do a lot of heat maps based on ZIP Code analysis. And so that would definitely be available.

DR. BURSTIN: And certainly, in our experience in the SDS trial, a number of measures have come forward, including those supported from CMS, that looked at SDS by ZIP Code, income, or even nine-digit ZIP, linked to the current

population survey at CPS. So, I think there are options that may get closer to the income issue, I think which is what Vanita's raising. It's really more income that might be broader than the duals.

MEMBER PINDOLIA: Yes.

DR. WINTERSTEIN: I think there's also something said to, in the discharge planning process, to think about affordability for patients.

To discharge a patient on a drug that clearly cannot be afforded may not really be the best course of action for that facility, and it might be better to find a lower-tier antipsychotic that would be affordable. So, I think there's also this -- there's a certain component that's in the control of the IPF, I would think.

CO-CHAIR PINCUS: Provide some incentive for them to consider the patient's particular situation. Okay. Andrew and then, Peter. And Les, is yours up? Oh, okay. Andrew

and then Les and then Peter.

MEMBER SPERLING: Okay. So --

CO-CHAIR PINCUS: You've got to make sure it faces us.

MEMBER SPERLING: My understanding, and I've worked in the Part D program for a long time, is the diagnosis for which the prescription is being written, or the indication is actually not part of the Part D claims data. So, the pharmacist doesn't actually know what the diagnosis is or even the reason the prescriber has written the prescription, right?

DR. CAMPBELL: Yes, that's correct. In our case, we've integrated the A and B and D data, so the diagnosis comes from the A and B data --

MEMBER SPERLING: Okay.

DR. CAMPBELL: -- for Medicare fee-for-service patients.

MEMBER SPERLING: And then another reason for the exclusion of dementia and these other indications is, because on the other side

of this, the Part D plans and the Medicare Advantage plans are all being held to account by CMS for off-label prescribing for dementia, and they are actually under obligation to lower that rate, and they've done so dramatically.

So you have to understand that CMS is using this for different purposes, for the Star Ratings, for these plans to make sure that they're -- and for the facilities to make sure that they are trying to reduce off-label prescribing of antipsychotics.

CO-CHAIR PINCUS: Les, and then Peter.

MEMBER ZUN: I'm always concerned about using face validity as your tool to measure validity. And when I look back at who assessed it, they were TEP, technical expert panel, but it doesn't say who was on the panel, whether they're physicians, what their background is.

CO-CHAIR PINCUS: No, it's --

MEMBER ZUN: It's really hard for me to understand who validated it if I don't know who the ten people were.

CO-CHAIR PINCUS: Yes, it's actually, if you get to the bottom of the --

MEMBER ZUN: Is it?

CO-CHAIR PINCUS: -- report, it actually lists the people. Yes, it's at the very, really, actually the very end almost.

DR. CAMPBELL: Yes, so, just to answer your question broadly, our technical expert panels are usually very diverse in terms of stakeholders. So we have folks that are clinicians. We have folks that represent hospital systems, folks that are methodologists, and also, importantly, patient and caregiver representatives on our panel.

CO-CHAIR PINCUS: Okay.

CO-CHAIR BRISS: I was going to say on the question of whether or not you should adjust, so at this point, I don't see good evidence that this should be an adjusted measure. I would call it significantly under the control of the provider. That if the provider does the right thing, that you can likely get a reasonable score

on this measure.

And I think we're way upstream about worrying about how and whether this might impact somebody's 5-Star rating, and I'm more -- at this point, you don't tell us anything, at least that I saw, about -- you saw a statistically significant difference between black patients and other patients, but we don't actually know how big that effect is. And so for a variety of reasons, I think it's way premature, actually, to be talking about adjusting this measure.

CO-CHAIR PINCUS: Other comments with regard to validity? So seeing none, let's vote.

MS. QUINNONEZ: Voting is now open for the validity of Measure 3205: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient. For the validity of Measure 3205: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

Looking for one more vote, if you could just submit your votes one more time, please? There we are, we captured it. All votes

are in and voting is now closed.

For the validity of Measure 3205: 9 percent voted high, 2 individuals; 78 percent voted moderate, 18 individuals; 3 individuals, which would be 13 percent, voted low; and 0 percent insufficient. So for validity of Measure 3205, this passes the validity criterion.

CO-CHAIR PINCUS: Okay. Let's move on to feasibility. One of the lead reviewers want to comment on feasibility?

MEMBER PINDOLIA: I think this one's an easy one. All the data is claims-based, and so it's easy to gather.

CO-CHAIR PINCUS: Any other comments about feasibility? Okay. Ready to vote on feasibility.

MS. QUINNONEZ: Voting is now open for the feasibility of Measure 3205: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient. For the feasibility of Measure 3205: option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

Looking for one more vote. There we are. All votes are in and voting is now closed. For the feasibility of Measure 3205: 61 percent voted high, 14 individual votes; 39 percent voted moderate, 9 individual votes; 0 percent voted for low; and 0 percent insufficient. For the feasibility of Measure 3205, this passes the feasibility criterion.

CO-CHAIR PINCUS: Okay. And now let's move on to usability and use. One of the reviewers want to comment on usability and use? David?

MEMBER PATING: I'm not going to comment, but I'd like to raise a question. I'm not sure who the target of this measure is. Is this a measure that looks at integrated health plans, or is it measuring --

CO-CHAIR PINCUS: It's hospitals.

MEMBER PATING: Hospitals?

CO-CHAIR PINCUS: Hospitals, it's part of -- that's --

MEMBER PATING: It's a hospital

measure?

CO-CHAIR PINCUS: It's a hospital measure.

MEMBER PATING: So on that basis, I actually have concerns about the usability of this measure. I think any measure that sort of measures the after-hospital performance, even if the prescription is given at the time of discharge, really discriminates against public hospitals with very transient populations that come in and out, are not identified, have no medical home, and most of the care is done in the hospital or sometimes even in a jail psych setting and then discharged.

So I'm really concerned about that. Even though the data is there, and it's easy to capture, these may not be linked patients, and so again, public sector hospitals, I would imagine will do poorly on this. VA hospitals, for that matter, maybe.

MEMBER JENSEN: VA hospitals, we would do great because we provide medications to our

patients. So it would be no problem for us.

CO-CHAIR PINCUS: And would the medications you dispense be captured?

MEMBER JENSEN: In this measure? No, because they're not measuring veterans. But I mean --

CO-CHAIR PINCUS: Right.

MEMBER JENSEN: -- we collect data about this all the time, about whether or not our patients pick up their prescriptions. And when they're discharged, we make sure they do, because the pharmacy is right there, and so the nurse goes down and gets it and sends the patient home with it.

MEMBER PATING: That would be for the two-day, right, and the seven-day. Just the 30-day, with homeless vets, I know sometimes are lost to follow-up. So that's -- anyway, the veterans system is not on -- because I don't think the Medicare is applying to that population.

CO-CHAIR PINCUS: But anyway, Rhonda

and Mike.

MEMBER ROBINSON BEALE: I would agree with David. I have some concerns for hospitals, not just public hospitals but hospitals that are not involved in some type of accountable care organization, because there's not necessarily those resources built in or those processes being built in.

So, if this would include populations that are not in ACOs, I think it's going to be problematic. Or Stars, if this is going to be a star and something that has to do with a rating in that regard, it will be problematic.

CO-CHAIR PINCUS: Okay. Mike, and then Peter.

MEMBER LARDIERI: Yes, I was just going to echo the same thing. I mean, I think, on a system level, I think this would be good, but just putting that on the hospital, when the hospital loses the patient, I mean, in our system, we don't have all our patients staying in-system -- they go to other systems -- how do

we do that follow-up with them?

I think at a system-level, you can hand-off from your inpatient to your outpatient provider within your system. They can reach out and have a push/pull type thing to make sure the patient takes their medication. I think that would be a better view of it than just on the hospital.

CO-CHAIR PINCUS: Peter?

CO-CHAIR BRISS: I'd like to speak for it, actually. So I think there's little -- everybody knows that there's a lot of losing people between hospital discharge and community care, but if we're not measuring it and paying attention to it and trying to do better, we're never going to do better, right?

And I work at a safety-net hospital; my clinical work's at a safety-net hospital. So I actually think this kind of measurement is essential to do.

CO-CHAIR PINCUS: Just for me to step out of the role of the chair for a minute, I

agree very much with Peter, because in some ways this is a lot more controllable than, for example, follow-up after hospitalization --

CO-CHAIR BRISS: Yes.

CO-CHAIR PINCUS: -- in seven and 30 days. Because the hospital can actually prescribe it, so it is a more controllable measure. Mike?

MEMBER TRANGLE: I mean, I agree, too. I mean, to some extent, what it's doing is, it's driving, even if it doesn't directly mandate, hospitals to use their outpatient pharmacy to have the patient leave with the script. That's the way you can control it.

So in some sense, I think it's kind of driving a practice change in a way that's not necessarily clearly articulated, but for you to do well at that, you need to take control and control what you can and make it happen.

Some people could argue that it's a cost-shifting for hospitals that are safety-net hospitals. Some of our hospitals are safety-net

hospitals. We've sort of bitten that bullet and sort of said, yes, we're going to do that. But I think that's sort of calling a nickel a nickel or whatever.

CO-CHAIR PINCUS: Okay. Lisa?

MEMBER JENSEN: Well, I think the other thing it's measuring is the organization's ability to be educating the patients about taking their medications. Is that happening during the -- at the beginning of their hospital stay right through to discharge.

CO-CHAIR PINCUS: Mike?

MEMBER TRANGLE: I'm really sort of probably getting ahead of ourselves for usability, but I do think the one caveat we sort of keeping come back to is it's a small slice of the population.

And I think, when we come to -- in some sense, when we come to voting for the overall thing, I'd like us to see if there's a consensus when we're doing that to vote with a recommendation to broaden the pool to include MA

patients or other ones that we could.

CO-CHAIR PINCUS: Okay. Other comments? So, let's vote on usability and use.

MS. QUINNONEZ: Voting is now open for the usability and use of Measure 3205: option 1, high; option 2, moderate; option 3, low; option 4, insufficient information.

All votes are in and voting is now closed. For the usability and use of Measure 3205: 22 percent, 5 individuals voted high; 65 percent voted moderate, 15 individual votes; and 3 individual votes, which would be 13 percent, voted low. 0 for insufficient information. For usability and use, Measure 3205 passes this criterion.

CO-CHAIR PINCUS: Okay. So now we are voting -- are there any -- there are no related or competing measures. So, now, we're voting for endorsement. Any final comments with regard to endorsement?

CO-CHAIR BRISS: I think, related to Mike's comment, that we'd already committed to

give that advice to CMS about we'd love to see a broader measure. So it --

CO-CHAIR PINCUS: Right.

CO-CHAIR BRISS: -- seems to me that it's okay to vote on.

CO-CHAIR PINCUS: That advice is thus given, Ellen. Yes. Okay. Seeing no further comments, let's vote.

MS. QUINNONEZ: Voting is now open for overall suitability for endorsement of Measure 3205: option 1, yes; option 2, no. Looking for one more vote. All votes are in and voting is now closed.

For the overall suitability for endorsement of Measure 3205: 87 percent voted yes, 20 individuals; 13 percent voted no, 3 individual votes. For the overall suitability for endorsement of Measure 3205, this passes this criterion. This is a recommendation.

CO-CHAIR PINCUS: So thank you. So, before we take a relatively short break for lunch, do we just want to hear from any NQF

member or public responses or comments to our morning's work? Anybody in the room? Anybody on the phone?

OPERATOR: To make a public comment, please press star-1. And we have no comments.

CO-CHAIR PINCUS: Okay. So why don't we take a break for lunch and let's return at a quarter of. And --

DR. LUSTIG: Yes, so we'll try to get your lunches quickly and we'll get right back to your second measure, if that's all right. Thank you.

(Whereupon, the above-entitled matter went off the record at 12:26 p.m. and resumed at 12:47 p.m.)

DR. CAMPBELL: Okay, so this measure, the medication reconciliation at admission was also developed for use in the same program, the inpatient psychiatric facility reporting program.

We know that inaccurate medication reconciliation is a significant patient safety issue that's been recognized by the Joint

Commission, Institute of Medicine, and other national organizations that have been focused on patient safety.

We also know that medication errors or discrepancies can lead to adverse drug events for patients and that a medication reconciliation process can reduce both the risk of errors and also harm to the patient through adverse drug events.

The medication reconciliation measures currently endorsed by NQF do not really assess the quality of the process, and they also do not focus on medication reconciliation at admission.

And in our review of the literature, medication reconciliation at admission was identified as a critical point in the care transition and also the point at which many errors could be introduced into a patient's medication regimen.

In our patient and caregiver interviews that we conducted, we had several patients indicate that a comprehensive

information gathering process was critical because they may not always be able to reliably report all of their information at admission when they're in crisis.

And they noted that they would prefer to go to facilities with overall good medication management practices, obviously including medication reconciliation process. And importantly, they felt that this measure was understandable to them and would be of benefit.

The measure addresses the quality of the medication reconciliation process by a establishing a minimum standard for the collection and documentation of information on medications that patients were taking prior to the inpatient admission.

It's specified as a composite measure to improve interpretation and has three components. The first is an information gathering component. The second is the completeness of critical prior-to-admission medication information, and the third is the

actual reconciliation step by a licensed prescriber meaning that you should continue, discontinue, or adjust.

We identified very wide variation in performance on this measure, suggesting ample opportunity for improvement. And it is a chart-abstracted measure. We did attempt to estimate burden, which we identified as a median time of approximately ten minutes per admission.

And we validated the measure with a random sample of approximately 100 admissions per facility. So we thank you for your consideration of this measure and look forward to your questions and comments.

DR. LUSTIG: If I could just jump in, I also wanted to give one word on why we're considering this as a composite measure. And I know different people often have different definitions.

According to NQF's definition, we consider this to be a composite measure because there are three components to the measure that

are aggregated for each component and then aggregated again for the final score.

And so for that reason, we did discuss with the developers classifying this as a composite measure. It actually required them to provide even more information than if it was classified as a process measure.

So we just wanted to clarify that in case anyone had questions.

CO-CHAIR BRISS: So with that, would any of the discussants like to tee up the importance to measure and report?

MEMBER PINDOLIA: Before we do that, I have one question, and I had emailed it to the NQF, and we wanted to have clarification. The measure is right now very general to say average completeness of med rec, processed within 48 hours of admission to an inpatient facility.

But everything inside, whether that's a gap survey, the reliability to everything, was based off nine psychiatric facilities only. So is this for just psychiatric hospitals, or is

this for all hospitals?

DR. CAMPBELL: Yes, it's for just inpatient psychiatric facilities.

MEMBER PINDOLIA: Okay, thank you.

MEMBER SUSMAN: It would help if you just briefly discuss the three components. And having not been one of the people who reviewed this in depth, I'm trying to get my mind around how you exactly compute these.

DR. WINTERSTEIN: Yes, thank you for this question. And that was quite a challenge because we really wanted to capture, via quality, medication reconciliation process rather than just presence of one.

So there's three components. One essentially looks at the sources that are considered in establishing a comprehensive medication, prior to admission medication list, because we are focusing on the admission process, so the key part is to find out what the patient took before he or she was admitted to the facility.

So we are looking in the first component at the sources that were considered, so the measure requires two sources. One that would be related to the patient. So that could be the patient him or herself as well as caregivers, pill bottles that were brought, or a medication list, and then a healthcare-related source which could be an outpatient prescriber, an outpatient pharmacy, the PDMP, so the Prescription Drug Monitoring Program, or any other healthcare provider that was contacted in order to get additional information.

In component 1, we also require that the medication -- this prior-to-admission medication list is in the designated place in the charts for easy reference and that the information that is on the PTA medication list is consistent with the information that might be dictated on the HMP by the admitting prescriber.

And when I say consistent, we're requiring that the PTA medication list includes, at a minimum, the medications that were also

dictated on the HMP. So that's the first component.

MEMBER SUSMAN: So this is all under one component, or is this --

DR. WINTERSTEIN: So each of the components, the subcomponents that I just mentioned would be scored, and that would be averaged across record. So basically there is a maximum of four subcomponents that would be met, and that would be your average per record.

A second component looks at the medications that are actually documented on the PTA medication list, and it requires that the name, the dose, the route, the frequency, and the last time taken is documented for each medication.

For the last time taken, there is permission to document that the patient cannot remember it because the patient is the only source that can reliably produce this information.

For the other four sources, there

would be a requirement that there's adequate documentation so that the reconciling physician actually knows what he or she is actually reconciling. That's the second component.

And that is again scored for each medication that is listed on the PTA medication list and then averaged across all medications that were recorded for the number of patents that were, for the number of records that were reviewed.

And then the third component is a single item, and that looks at whether the medications on the PTA medication list were reconciled by a prescriber within 48 hours. And the reconciliation is operationalized with either documentation that the medication was to be discontinued, discontinued, or modified.

So there should be a prescribing decision that specifies what is to be done with this medication, and that should happen within 48 hours. And that's again average for all medications that are extracted across all

records.

MEMBER JENSEN: Okay, so one of our first comments from the folks that reviewed this was that the literature is weak on supporting this. And so I'm wondering of you can talk a bit more about the literature and what kind of support you all found in developing this.

DR. WINTERSTEIN: Yes, so the literature is, and I think we all are familiar with the IOM report in 1999 that came out on medication errors and medical error in general and the prevalence of preventable adverse drug events in both institutional and non-institutional settings.

So we know that as a whole, preventable adverse drug events are quite common and are considered a very massive threat to appropriate healthcare and optimal healthcare outcomes.

That said, and I actually personally have done a lot of research in this area, there's also a huge challenge in measuring this because

preventable adverse drug events come in many flavors and forms.

And so to measure them really requires systematic chart reviews by very astute people who are actually able to identify preventable adverse drug events as such. So it's not so easy to measure this, and that I think explains the major limitations for most studies that have been published in this area.

There are a lot of studies that have been published with respect to medication errors on admission. Usually these are studies that have been conducted by pharmacists that look very specifically on medications that were ordered on admission and compare those against what patients were admitted on.

And the range of problems, ranges, depending on the definition really from 10, 30 percent up to 80 percent per patient who is affected by at least one error that could have been avoided.

But are those errors then really

relate to sincere outcomes, as I said, is a very different question. I think in most studies, there's an estimate that about 10 to 25 percent of those would be considered clinically-relevant in one way or the other.

In many instances it's omission errors and admission errors so that essentially, essential medications were forgotten and not continued on admission. But there also are, you know, confusions, different doses, and so on.

CO-CHAIR BRISS: Other comments on, or questions on importance of the measure?

CO-CHAIR PINCUS: Just one question. Sorry.

CO-CHAIR BRISS: Andrew.

MEMBER SPERLING: Thank you. So my question, or really comments relate to the accuracy of the data that's going to come out of this. I'm specifically concerned about instances of acute psychosis, suicidality, really the extreme symptoms.

And that's really the cohort of people

we're dealing with, right? People that are adhering to their treatment and are doing fine aren't admitted to psychiatric hospitals, because we know there's aggressive utilization management on the front end in many healthcare plans.

And the question is, the ability of someone who's in the midst of an episode of acute psychosis to render an accurate list of all the medications they have, not just the psychotropic medication, but if they have comorbid diabetes or asthma, a whole long list, to be able to come forth with the names of the medications, the dosage, the frequency, all of that. So we're collecting data that clinicians are going to rely on that may not be entirely accurate.

And so that's one of the concerns that many of us have of making sure that, if we're collecting data, let's make sure it's accurate so it's used in the right way.

And I know you've made some effort to look at when a patient being admitted simply can't deliver that list of medications and how --

can you talk a little bit more about how you sort of accommodate for those emergent circumstances, which is, quite frankly, more prevalent in the case of an inpatient admission.

DR. WINTERSTEIN: Yes. And we did alpha test this measure when we started to really define those specifications in our own two inpatient psych facilities. And we got a lot of feedback from those providers there.

So if the patient is unable to provide information that can be documented, and if that documentation is there, that would be counted as an appropriate patient source. So actually the facility would score favorably if -- as long as the facility documents this patient was not in a condition that he or she could provide any useful information.

So that's absolutely accepted as a patient source. In order to address the issue that this is a really very special circumstance where many patients may really not be able to provide their accurate PTA medication, this was

the exact reason why we require a health system source as well.

And that's what we consider the minimum standard, that there's at least an attempt made to try to get a more comprehensive view on the prior-to-admission medication list.

MEMBER SPERLING: And what would those be if there isn't a family member there? I mean, they have to randomly call pharmacies, or what?

DR. WINTERSTEIN: Yes, so the health system sources would be an outpatient provider. So any kind of PCP who could be identified, outpatient pharmacies, the Prescription Drug Monitoring Program that is accessible to all providers in 49 states.

Prior to admission, prior admission records or records that would be available within the health system that the patient was admitted to. All of those would be acceptable.

MEMBER SPERLING: You mentioned --

DR. WINTERSTEIN: Just one of those.

MEMBER SPERLING: And this covers

psych units and general hospitals as well?

DR. WINTERSTEIN: The measure is only constructed for inpatient psychiatric facilities. It was only tested there. But if those have an integrated electronic health record system and can look at the affiliated hospitals or affiliated clinics, they could use that information as well.

CO-CHAIR BRISS: I think I have Jeff, Harold, Michael. Jeff?

MEMBER SUSMAN: So I'm interested in the evidence. I mean, so IOM, whoever comes out and says, yes, we should do medicine reconciliation, but this measure strikes me as very complicated and taking a lot of things together in a composite where I'm not sure that any element is equal.

I mean, is it really just as important to do A as the third component, composite 3? I'm not clear that there's good evidence that doing this in the way that it's constructed and measured here leads to improved outcomes

although, yes, you can say it feels good.

But does it really make a difference? And I guess I'm somewhat skeptical. We can talk about other matters which I'm concerned about, but let's just start with evidence.

DR. WINTERSTEIN: Yes, I think that was one of the major points to reading arguments to really submit this as a composite measure to give you the opportunity to look at the contribution of each component to the overall score at the ends and to make it very transparent that there are three components.

And we had a lot of discussions, and you will see in the sensitivity testing that was done that there's actually various ways of summarizing the components.

There was one way that we present where we ignore them altogether and just take all the elements for the measure and summarize everything up. What we ended up presenting here, and you will see the data on how this compares for each of the various versions, is that there

was a 30/30/30.

So component 1 is the information gathering process. Component 2 is here's the meds, and here's the information on the medications. And number 3 is the medication reconciliation.

Number 3 is just one single item, but that's the action step that is the finite really trigger for an outcome, right? So if the physician doesn't reconcile whatever is on this medication list, nothing is going to happen.

So that last component is obviously very critical. However, this last component also just only works when there is actually information that is valid that an action can be based on.

So as we were discussing this, to us and to the -- we actually had two groups. One was an internal workgroup of providers in our own facilities, and then we had a larger technical expert panel that is again in the materials listed.

To us, it felt -- to that whole group, it felt that the way we had assigned value or rate makes most sense thinking about how each process contributes to the overall effect at the end, which would be hopefully we have a medication reconciliation process that ends up in the right medications.

MEMBER SUSMAN: So just to briefly follow up, are there studies which would link each of these components, or the elements that go into the components, with enhanced outcomes. And that's what I'm trying to get my head around.

DR. WINTERSTEIN: Yes. So those studies that have -- well there's two things. One is the components that we mention are aligned with the process that the Joint Commission defines as an appropriate medication reconciliation process. So they are consistent with that.

MEMBER SUSMAN: So it's not an evidence-based recommendation?

DR. WINTERSTEIN: Right. No, that's

just a best practice. Those studies that have been published, that have looked at a medication reconciliation process, use that process. In most instances, they are pharmacists, and most studies have been conducted in acute care facilities and not in IPFs.

In most instances, they are pharmacists that use an information gathering process that is two-pronged, similar to what we are proposing, that ensures that the medications are properly recorded. And then there is a physician reconciliation process at the front end.

So that is kind of a standard process that we were trying to capture, yes.

CO-CHAIR BRISS: Harold and then Michael.

CO-CHAIR PINCUS: So I had several questions to clarify exactly what the measure is in terms of its operationalization. And then I had some concerns.

One is, how are you defining attempt

to reach? Because it doesn't say that they necessarily reach a source. So if you could talk a little bit about that.

And also to clarify, this includes all medications, not just psychiatric medications?

DR. WINTERSTEIN: Yes, so we define as an attempt, as a documentation in the chart, that an attempt was made to contact the provider.

CO-CHAIR PINCUS: So what does documentation of an attempt mean?

DR. WINTERSTEIN: I called the pharmacy, and the pharmacy was closed, or I called the provider, and the provider was closed, something along these lines.

CO-CHAIR PINCUS: What if there is no known provider or no known pharmacy?

DR. WINTERSTEIN: There would still be a PDMP that could be contacted at the minimum, as long as there was a patient name.

CO-CHAIR PINCUS: What if there was no PDMP, if the patient has no insurance?

DR. WINTERSTEIN: Well, there would

still be a PDMP. The PDMP is not linked to insurance. The PDMP would capture everything that a pharmacy dispensed for this patient that is a controlled substance. It's not linked.

CO-CHAIR PINCUS: As a controlled substance?

DR. WINTERSTEIN: Yes. Yes. PDMP is Prescription Drug Monitoring Program.

CO-CHAIR PINCUS: Right, right. But that's only for a controlled substances.

DR. WINTERSTEIN: Yes, that's correct.

CO-CHAIR PINCUS: So it's not for anything else.

DR. WINTERSTEIN: Yes, yes.

CO-CHAIR PINCUS: So --

DR. WINTERSTEIN: So that was the first question. Remind me --

CO-CHAIR PINCUS: And the other is -- but this is medication reconciliation for all medications?

DR. WINTERSTEIN: Yes. Yes, and there's two reasons for this. Number one, a lot

of psychiatric patients have certainly a lot of non-psychiatric medications. There is certainly an increased risk for non-mental morbidity in those populations.

CO-CHAIR PINCUS: I know. I mean, I think it should be obviously for all medications. Yes, I'm not asking for justification for it. I just want to be clear that that is in fact --

DR. WINTERSTEIN: Yes.

CO-CHAIR PINCUS: Although for one of the sources, it's only for a single class of medications.

DR. WINTERSTEIN: Yes, that's correct.

CO-CHAIR PINCUS: So that you could actually meet that criterion just by looking just at a single class of medications.

DR. CAMPBELL: Yes. And I think it's important to note with the measure that we're attempting to establish a minimum standard of what might be necessary for a provider to do, and not necessarily what might be required every single time for a provider to do.

So while that might be sufficient in one case, in other cases, the provider may need to contact two or three different health systems to identify that, correct?

CO-CHAIR PINCUS: So here's my concern, is that this is very minimal. It's really very limited in terms of both the nature of the data collection, but also, it's just about documentation.

It doesn't look at all about the accuracy of the medication reconciliation, either accuracy in terms of whether the documentation is in fact accurate to what the patient actually has been receiving, nor does it look at the quality of the judgment about reconciliation.

So it's very minimal, and it's heavily burdensome. And so it raises the question of whether the juice is worth the squeeze here.

DR. CAMPBELL: So if you look at the -- I mean, I think if we go back to the discussion that we had on the prior measure, if you look at the data, there's a huge quality gap

here to suggest that the facilities aren't even yet doing the minimum.

And while we can't guarantee that what comes out of this process is 100 percent accurate medication list, what we can guarantee is that if they don't go through this process, there's evidence to suggest that discrepancies are going to occur.

So if the dose, the name of the medication, the frequency, all those types of things aren't documented in the medical record, and there isn't documentation in the medical record of an action step occurring, it's highly likely that there's going to be an error which then will lead to ADE.

I don't think at this point there's any feasible way to measure, that wouldn't be incredibly burdensome, the overall validity of a medication list at the end of the day. But I think the process that we're suggesting here would greatly reduce the chance that a discrepancy would occur.

MEMBER TRANGLE: I think this measure was so complicated it's hard to know where to start to make a comment. And maybe I'll start, as most people do, from sort of the ones that are sort of engaging your emotions more than your intellect, you know?

In our system, we've been working on this for a while, in Minnesota, and we have an integrated system even though 13, only 13 percent of our safety net hospital admissions are on our health plan. But we can directly see it. And a slightly higher percent are on our EMR, which we share with our 1,800 outpatient docs.

And we've been working on this, and it just feels like we're torturing ourselves to some extent. I mean, I think all of us buy into the concept that if you know what somebody's taking, you can coherently make decisions which you have to make, and it's much, much better for safety and quality of care.

So the concept, the need, that the importance of it is sort of, and at least in our

system, is totally a given. How we've struggled to operationalize it has been sort of quite variable. And I think our results have been very variable.

Theoretically, we have pharmacists in our ED some of the time, and they'll try to do it. But it seems like more than half the time they'll say: tried to connect with something and couldn't do it, which if you're the attending the next day, figuring out, well, what am I going to order, is totally useless. You know?

At the same time, we'll try to get a hold of the pharmacy or a person or someone else, and we try to do all this stuff, but it's -- and we have an EMR that drives us to either sort of continue, modify, or discontinue the med.

But in a lot of ways, documenting that, the documentation by itself has been quite onerous and hasn't necessarily helped us. What I really care about is what were they taking, what's our best guess, and how do I use that data to make coherent decisions and decide what to

start them on and what not.

And the documentation part doesn't really help, especially if it's documenting that we tried and didn't learn anything. But it's quite onerous.

And I guess where I'm coming form is sort of like I totally think it's a really important measure to figure out how to make it work and how to do it. But it needs to be simpler, more operational, and less time-consuming.

And that's my own personal, but I also know it's my system's experience. Did you get much feedback when you asked other places or people for feedback? Did you hear similar things, or what were the recommendations here?

DR. CAMPBELL: Yes, so I know this is getting more into usability. So we have heard public comment about the potential burden. And, one of the things we believe is that, given that the measure was validated in a limited sample of 100 cases, we believe that that burden could

potentially be limited if the measure were implemented.

The other thing that we've planned to do is to provide an electronic tool to facilities that don't have electronic health records that would allow them to calculate the measure directly.

DR. WINTERSTEIN: I think your question was more about doing what the measure requires rather than extracting the measure, right?

MEMBER TRANGLE: We could see the burden, maybe more -- we have -- I think I'm on now. We have an EMR. But looking at the EMR doesn't give us any guarantee that it's accurate. You know? A lot of meds are on there, and they're on there forever. And unless someone is sort of updating it and culling it, it's not necessarily reliable either.

DR. WINTERSTEIN: We have the same issue in our facility, yes. I think there is -- you're coming from a level that is already

somewhere here on the ceiling. Just based on our sample of field testing hospitals, we have some facilities that are pretty much on the street, just in terms of --

Yes, I mean, and paper is not necessarily bad for a good PTA medication list. But we have facilities who don't have even a paper form. We have facilities where there is no designated form of any PTA medication list whatsoever.

MEMBER TRANGLE: I didn't really mean on paper. I meant someone, a pharmacist said: tried to contract somebody. Or: tried to update this and couldn't.

DR. WINTERSTEIN: Yes. So let me explain why we required in this measure this documentation of two sources. And that really goes back to what was mentioned earlier about the lack of this measure to evaluate whether the medication, whether the PTA med list was right.

And the main reason why this is very difficult to do is because there's no gold

standard of what is right. Right? I mean, even if you call the outpatient PCP, he knows half of the pie, and any specialist will know the other half of the pie, and the pharmacy will know a certain piece. And the patient is the only one who can tell us whether he's actually taking what he even filled.

So there is no gold standard. It is a process, and there is no outcome. There is no accurate PTA medication list. So what we were trying to do is to essentially define a process that would increase the likelihood that there's actually a workable PTA medication list that a physician could use.

Will that be the optimal and perfect and accurate list? Absolutely not. But really it would be better than what is currently available, which in many of those facilities is zero. I would argue yes.

And I completely agree with you that to get this process absolutely right, it is extremely difficult. We have the same thing in

our facility. But from here to there, there's a lot of other things that can be done.

CO-CHAIR BRISS: So we may be jumping around a little between criteria. So we've been back and forth a little from importance to feasibility to usefulness. So we might want to try to do one criterion at a time.

I agree that it's, on this one it's a little hard. And so I want to take off my chair's hat for a second and make a couple of comments, and then we'll just go the rest of the way around the table.

So I think I also have some concerns about this one. It's specified in a way that's complicated. It's hard to understand. I won't talk too specifically about burden here, but it's a little hard to know how much upside I'm going to, I'm likely to get from implementing this measure relative to what it costs me to implement it.

And so, and the systematic review doesn't help me much with that. It mostly reads:

we looked at 25 or 30 studies; most of them were not any good. The ones that were any good mostly showed improvement in charting and not actually improvements for patients.

And so it's a little hard to know whether this very complicated burdensome measure is giving me a meaningful upside that makes up for those things. And so I think I have some concerns.

And then having said that, it's inarguable that this is an important problem. But it's -- whether this specific measure is the right way to solve it, I'm not sure.

And so with that, my chair's hat comes back on. I would like to move us relatively quickly to a vote. And so I may have contributed to the problem that I'm about to identify.

But I think that we're doing some degree of repeating each other. And so try to be relatively brief, and try to, if you feel the need to agree with what's already been said, maybe note that. But let's try to get through

the rest of the comments and get to a vote on this criteria, please?

MEMBER PINDOLIA: So I'll be really brief. I guess I'm saying that I agree with the comments that have been made. And the work we've been doing at Henry Ford Health System with trying to improve the med rec upon admission I feel would go backwards if we added another burden of here's a checkbox, because it's going to be another checkbox that we'll get sidelined from the work we're already doing to try to improve that with our limited resources.

MEMBER ROBINSON BEALE: Yes, I agree it's very complex, but I also agree it's very, very important. So I'm wondering whether or not the focus of this being on the hospital may be an area that is later in the process.

And maybe what it needs to be is a focus on the health plan who has the utilization data. The hospital generally calls to let -- to find out if a patient is someone with insurance. The hospital's got to call to find out if the

person is eligible or their insurance is active.

So they notify the health plan of this admission. The health plan in most cases will have access to pharmacy data, at least the fill rate. And so I would see this, and I will deny this outside of this room even though this is going on record, but the process should be on the health plan to make sure that information is available to the hospitals for any acute admission.

I think you would have a better bang for your buck if it was placed there and not on the hospital. So I'll stop right there.

CO-CHAIR BRISS: We can get to a vote.

MEMBER MAZON JEFFERS: I was going to say something similar, that it's almost as if the three steps are additive to the final step. As you've said yourself, you can't do the reconciliation if the information isn't good.

But maybe how you go about doing that is less of a concern than actually achieving the, maybe, and it's not perfect but if there's a way

to focus a little bit more energy on the reconciliation piece as the actual measure and not all three steps, and allow systems to figure out what they need to do in order to achieve the reconciliation piece which really you have to do those steps anyway in order to get there.

MEMBER LARDIERI: Yes, and so thank you. My question is how is this different than under meaningful use for hospital, have to do medication reconciliation, or you just figuring them no psychiatric hospitals have EHRs, some do and they would already have to have this incorporated in. So how is this different than that?

DR. CAMPBELL: I'm sorry, so how is it different than the meaningful use type measures for electronic health record medication?

MEMBER LARDIERI: For the medication reconciliation. There's a measure specifically that they have to meet for medication reconciliation.

DR. CAMPBELL: But those types of

measures are more type of attestation which again when we talk about checkbox type things, that's what we're trying to get away from. And so what we were trying to do with this measure is to operationalize something that's meaningful in terms of understanding the quality and the completeness of the process.

So I think there's a balance here that if you don't try to accurately measure the process itself and you just simply say the process was done, then it could become a checkbox measure, right, that everybody just said yes, the process was done but we have no idea what the quality of it is.

And I think what we're trying to achieve with this measure is that at a minimum, if you don't have, like, dose frequency and that sort of thing about the medications for the patient, and you don't know whether the patient should be continued or discontinued on a medication within 48 hours, that's not a hospital that I would want to personally go to. Right?

(Simultaneous speaking.)

MEMBER LARDIERI: The measure under meaningful use is at, you have to identify dose, frequency, the same thing. So it's not any different. It's not just a checkbox, you have to identify the dose and everything.

DR. CAMPBELL: Okay. So I mean, my understanding of the measure was that it was more an at a station measure. So I guess that's something we need to look into.

MEMBER LARDIERI: I want to review it because that's not what it says. The other piece is, and I heard that you're going to give a free web-based platform for organizations that don't have an EHR. But if they do have an EHR, have you estimated the cost to make this form because you're requesting a for in the EHR.

So that means development time, that means a cost to the provider to pay for my vendor to do that. So what's that overall cost across the country?

DR. CAMPBELL: Yes, so we haven't

looked specifically because so few inpatient psychiatric facilities have implemented electronic health records at this point. And at this point, the measure is intended only for inpatient psychiatric facilities. So it hasn't been specified as an EHR measure.

MEMBER LARDIERI: Yes. And you know, I like the idea of the web-based platform, but if we're going to do a web-based platform, give it to them for everything they need to measure, not just for this one thing would be helpful.

CO-CHAIR BRISS: So we need to vote on importance of the measure and reporting. So this one's a bit of a tough one. So we're trying to reframe the question, right? This is supposed to be just on the evidence, right?

But I let the conversation go a little broader because I think that you can't answer the evidence question without knowing sort of what the quality measure is and what the intervention is.

And you may not be able to vote on the

evidence question without making some judgment in your head of, does the upside, is the upside worth the potential downside in terms of costs or harms or other things.

And so I let it go. But the question on the table is is there evidence about that if we do this, essentially if we do this measure, that it's going to be a measurable upside.

CO-CHAIR PINCUS: You just summarized what I was going to ask. So that was fine.

CO-CHAIR BRISS: We've done this together a lot. We can probably channel each other. We're like an old married couple. And so with that, can we vote on the importance? It's measure and report.

MS. QUINNONEZ: We are now voting on the evidence of Measure 3207, Medication Reconciliation on Admission. Option number one, high. Option number two, moderate. Option number three, low. And option number four, insufficient.

CO-CHAIR BRISS: And just in case this

comes up, can somebody remind the panel the difference between low and insufficient?

DR. LUSTIG: Insufficient really means you don't have enough information to make a decision.

MEMBER MAZON JEFFERS: Can we just be clear then that if you don't, if people vote low then the measure won't pass and it's a must pass criteria.

MS. QUINNONEZ: Voting is still open for Measure 3207 for evidence. Option one, high. Option two, moderate. Option three, low. And option four, insufficient. All votes are in, voting is now closed.

For evidence of Measure 3207, 4 percent voted high, one individual, 26 percent voted for moderate, 6 individual votes, 65 percent voted low which is 15 individual votes, and one voted for insufficient.

CO-CHAIR BRISS: Okay. And I just want to emphasize to the developer that we know that this is a lot of work, right, and nobody

doubts the importance of the topic. But there may be some, I hope you find something to, compensation to be constructed and moving this work forward.

DR. CAMPBELL: Yes, absolutely. We appreciate --

(Simultaneous speaking.)

DR. LUSTIG: We actually should have just a little bit of feedback on specifically what type of evidence the Committee would want to see going forward and --

MS. JOHNSON: And I think it would help us when we write the report to understand since there was a lot of discussion about the quality construct and how it was created and feasibility and all that kind of stuff kind of got meshed into your vote.

We want to make sure that your vote was actually about evidence. So you, we want to make sure that you really felt like there's not enough evidence to show that medication req leads to good outcomes.

CO-CHAIR PINCUS: Yes. I think it was basically the point that Peter made at the very end. It was the link between the evidence that the documentation of this construct would actually have an impact on patient outcomes.

CO-CHAIR BRISS: He's channeled me again. Raquel?

MEMBER PATING: For giving comments for future suggestions, if you bring this back, I would not bring this back as a paper measure. I would try to move with a measure as complicated as this strictly into eMeasures. I think that the burden, the systems would be really hard.

And it wasn't quite clear to me, because I saw some notes saying this was a paper, I guess one of the National Hospital Association or Public Hospitals recommended eMeasure and I would recommend if you came back you look at only that format.

CO-CHAIR BRISS: All right, and one more thing. One of the things that I found hard in the materials is getting your head around what

the measure is exactly measuring and what the scores mean is really hard if you're coming to this cold.

And so it's sort of translating from that to if we do that, will you have better outcomes was very hard for me. And so maybe the staff has been giving me funny looks about aren't you creeping into reliability or validity or other things.

But I said already that it's hard to answer the, even the evidence question without being able to answer what are you measuring and how does it improve things. And I think you would be better off with a simpler measure, if you could do that.

DR. LUSTIG: Thank you.

MEMBER TRANGLE: I mean, it could be very concrete. If somehow, even if we started out with, like, .001 percent, if we were documenting these were the meds somebody was truly taking before, or what they were, you know, versus I made an attempt to find out, it would be

much more meaningful to me and I think that grade of evidence would be higher.

It means something to me and the rest would flow from that, the feasibility, the usability, and you know, versus we asked and documented that we asked but couldn't find out.

MEMBER SUSMAN: And to make sure that your lives are doubly complicated or triply complicated, the issue of it being a composite measure for me and tying it back to evidence was very difficult, and the reason I chose in part to vote for insufficient evidence or low evidence.

I think, you know, maybe the advice you had form NQF staff or the approach that you decided to take around making this a composite might have complicated this discussion and particularly just focusing on evidence.

CO-CHAIR BRISS: All right, thank you. We are really close to back on time.

CO-CHAIR PINCUS: Let's move to the next measure which is prevention, care, and screening, tobacco use screening, and cessation

intervention, 3225.

DR. LUSTIG: And so I've actually been remiss at announcing when there are people who have recused themselves from measures. So Dr. Briss is actually recused from this measure, the e-version and the prevalence measure that are coming up. So he won't participate in any of the discussions.

And once again, to add to your confusion, this is another one of those that this is the previous 0028 but has been renumbered because there is also the e-version that we will look at.

CO-CHAIR PINCUS: So if the measure developers might introduce themselves and present some preliminary information.

MS. TIERNEY: Yes, thank you. Hi, everyone. I'm Sam Tierney, I'm the Director of Measure Development Operations at the PCPI and I'm joined by my colleagues Jamie Jouza who is with our specifications team, and Eduardo Segovia who is with our testing team.

We appreciate the opportunity to be here today and to present this measure and the versions of it, the various versions of it for consideration for continued endorsement.

So just a little bit of background about the measure. The measures if you will. They are intended to promote adult screening for tobacco use and the provision of cessation, intervention for those who screen positive.

Although the impact and public health importance of tobacco screening and intervention is well founded and I'm sure very well known to the Committee, since our measure, the first tobacco measure to be reviewed, I thought I would provide just a little bit of highlights about the importance of this topic area.

Tobacco use is the leading preventable cause of disease, disability, and death in the United States, cigarette smoking results in close to half a million of premature deaths each year and accounts for one in every five deaths.

There is good evidence that tobacco

screening and brief interventions are successful in helping tobacco users quit, and that if people quit, that can lower their risk for heart disease, lung disease, and stroke.

The measure was developed originally many years ago. It has continued to be updated by us on an annual basis and a more rigorous update every three to four years.

The measure was developed by a multi-disciplinary cross-specialty workgroup. And it has undergone some important milestones in the measurement process. We have a rigorous process we follow including a public comment period, and we previously would submit our measures to vote by our membership.

The measure was developed to align with the USPSTF recommendations around tobacco screening and cessation intervention. And as a result it is focused on adults. I know there was a comment about potentially expanding the measure to adolescents.

There is a corresponding

recommendation statement from the USPSTF around screening or adolescents. And there are some unique needs for that patient population which are addressed by another measure that NCQA stewards that particularly deals with that patient population.

The measure certainly addresses a gap in care, and as you probably noted from looking at the forms in the PQS program, there's still a demonstrated opportunity for improvement.

Although the rates seem relatively stable over time, the adoption of the measure is fairly low with about 21 percent of providers reporting.

More data that's probably more nationally representative although probably a bit dated comes from a medical literature and we included it in our submission that indicates that rates of screening are around 65 percent and then rates for intervention for tobacco users are even lower, around 30 percent.

There's also data that indicates that

there's disparities in screening and intervention based on age, ethnicity, and insurance status.

The measure is in widespread use in various federal programs including the new MIPS programs including the new MIPS program, the former PQS program, and the EHR incentive program.

And we believe that continued endorsement will only continue to encourage its adoption, and continue to impact the lower rates that we've seen, the suboptimal rates of tobacco screening and cessation intervention. Thank you.

CO-CHAIR PINCUS: Do we want to move to the initial criterion? Connie?

MEMBER HORGAN: Just a comment, in your list of disparities was the standard sociodemographics. I just wanted to ask a question about disparities related to individuals who have mental health, alcohol, and drug problems in the literature in that area.

MS. TIERNEY: You know, that's a great question. I do know that it's, you know, it is

an important special population to look at as it relates to tobacco use screening and intervention.

I know actually Dr. Pincus, you authored an article where you talked about separating that category of patients out as sort of its own disparity element and identifying how the rates compare with that patient population as compared to others.

We did not provide that in the submission form. I would be certainly happy to, you know, review the literature and provide that, but I do know there is a version of this measure was adapted by the, by NCQA for use in patients with that sub-population, with serious mental illness.

So it is an NQF endorsed measure focused just on that, that more narrow population. Although this measure, because it's very broad in nature, does include that population, we don't stratify the information out by that variable.

CO-CHAIR PINCUS: Just so why don't we move ahead to actually looking at issues around sort of evidence. Do we need to look at that since this is a maintenance measure?

DR. LUSTIG: I think that there was some new evidence provided. Was there updated evidence?

MS. TIERNEY: Yes. USPSTF had done an updated recommendation statement from 2015, and they had an additional evidence review that supported the recommendation statement was all still favorable and in support of the measure focus, but it was updated since the last time it was reviewed.

CO-CHAIR PINCUS: So it sounds as if unless people have specific comments, we don't need to vote on that, correct?

DR. LUSTIG: Yes, because even though it's new evidence, it's directionally the same as what we had before. So unless there was objection, we don't need to re-vote on evidence.

CO-CHAIR PINCUS: Mike, are you

concerned?

MEMBER LARDIERI: I just had a question. Why 24 months? It seems if somebody starts smoking, and maybe I don't know the evidence around addiction with smoking. If I start smoking this year, does it take 24 months before I get addicted?

So I'm just wondering why 24 months. We found that not useful in our system, so we do at least once a year. So that's one question.

The other piece was around kids, and I guess it's around folding adolescents into this as opposed to having two separate measures because that's just more of a burden. But first on the 24 months.

MS. TIERNEY: Yes, that's a good question. So initially when the measure was developed, we sought the input of an expert workgroup in trying to determine sort of the periodicity of the requirement for the numerator.

And at that time, the group felt that 24 months was appropriate. I know that we did

look at the literature, and I don't believe there's any particular evidence that indicates an appropriate timeframe, whether that be six months, a year, 24 months. So I think that was what our expert workgroup agreed to.

And with the lack of particular literature that recommends how routine the assessment should be, that's what the measure includes.

I will say that the measure does include language about at least once every 24 months, and what we've heard from implementers is that oftentimes they do do it more often. Some do it at every visit, and some do it more on the annual basis.

So the measure doesn't prohibit someone from doing it more regularly, but the base requirement is that it be done once every 24 months.

CO-CHAIR PINCUS: Raquel?

MEMBER MAZON JEFFERS: I think I might be skipping out of order of the discussion, but

this measure is a lot like 0027 which is the next, not the next measure but it's, like, three measures away that we're going to be, I mean, a lot like it.

DR. LUSTIG: There are different levels of analysis, and 0027 is based on patient reported data.

CO-CHAIR PINCUS: Mike, anything specific about the evidence?

MEMBER TRANGLE: I think so. It's really to some -- there was an article that came out by Mike McKusick maybe three, four months ago talking about the best value of doing sort of interventions in terms of impact for what you invest in doing it.

And that article basically said the three, the top three things of anything, this is medical not just psych stuff, is smoking cessation work with adults, smoking cessation or prevention and cessation with adolescents, and then doing vaccinations.

So to the point that was made earlier

about the age range and, you know, is somebody looking at the adolescent group. And if it isn't, it might be something we might consider recommending this group work on in the future.

MS. TIERNEY: Yes, I think it's a great point. I do think the age cutoff is somewhat arbitrary. I think it results from the fact that when we were developing the measure, we were focused on preventative services and screenings for adults.

There is another measure that, so this notion is certainly covered within the NQF portfolio because there is another measure that includes patients between the age of 12 to 20 and includes, you know, is a very similar focusing on tobacco cessation and intervention.

I will say the one challenge, you know, with sort of combining it within the one measure is that there may be specific concerns or considerations for the pediatric population such as, for example, no pharmacotherapy or maybe focusing more on a second hand smoke exposure.

There may be particular recommendations for that population that if we were just to consider adding them to our measure, that sort of the nuance of that may be lost a little bit.

So I do know that there is, again, there is that other measure. But that's, but currently yes, our measure is just focused on the adult patient population.

It's certainly something we could take back to consider when we update the measure. But of course at that time we would have to look at again sort of those unique needs of that patient population as well as the measurement landscape in trying to not, you know, to avoid duplication and think about harmonization and that sort of thing.

But it's certainly something we could consider for the future.

CO-CHAIR PINCUS: Let me just say that, you know, what we've decided to do because we were getting sort of delayed is that time

permitting, to have a discussion about alignment across all the tobacco measures tomorrow after we've been through this.

And, you know, the staff at NQF had actually done, put together a nice table sort of looking at the different tobacco measures. But our thinking was let's go measure by measure and get through. Okay?

So is there any objection to moving beyond the evidence and now going to gap? Okay. So with regard to gap, do any of the people that have been involved in the initial review want to comment about the gap issue? You know, is the gap that was present when this measure was initially developed, is there still a gap.

MS. JOUZA: Yes, so performance gap was the comments by the group that looked at this was more focused on disparities and disparities from the public data, and that that seemed to not be directly addressed in the measure but there.

And I was, I myself was particularly, I don't think we have time to discuss it, but

interested in the -- that cessation assistance is higher in the Medicaid population by some published data, and those in high poverty areas as compared to the private and Medicare population which surprised me, and those in low poverty areas.

By age was still disparity. And in terms of the performance gap, in terms of reporting, I think the biggest gap, perceived gap is that even though there's good performance for those who report, only 21.7 percent are actually reporting. So there's still a lot of room for improvement there, at least as in measure and PQRS.

I don't know if someone else from the group wants to comment.

CO-CHAIR PINCUS: Any other comments about the performance gap?

MEMBER PINDOLIA: Couple questions. So looking at the 2015 PQRS, I mean, there was a tremendous improvement in, so the claims is it's at 96.24 percent. And I understand we're saying

only 22 percent are reporting.

So the comment that I have, there's been a huge movement with many employer groups and to do health risk assessments and for their employees.

So the patient comes in for a visit but the doctor's not documenting that this was a smoking cessation, but it gets documented on the MQF or on these other forms that are submitted. Could that be where we're missing so much that is going on, because there are so many employer groups that want that record.

And so for health plans, we have a lot of that data. But there is nothing that comes in that there was a coding for smoking cessation discussion or that because they filled out what they called an MQF form or other forms like that.

MS. JOUZA: So with respect to the way that this measure, and I know we're talking about the claims registry piece, the denominator coding that we would require in order to pull a patient in isn't such that it would require that it

states that there's any sort of smoking cessation, there are two CPT codes that are included right now, 99406 and 99407 which are specific to that cessation, but again, not required.

And the way that it's captured in the claims registry world right now for PQRS is is sort of that attestation with the CPT2 4O or 4F that says I screened my patient for tobacco use, and if they were a tobacco user, I intervened.

MEMBER PINDOLIA: Well, I can tell you from my experience at least just within Henry Ford Medical Group because when they have to do this, let's say for Ford Motor Company that's in our backyard and these large employer groups, it causes such a huge unevenness in their patient workload.

I mean, the documentation has been very streamlined. It's just this form, and this is what you're doing and this is what it is. So I don't believe they document that.

And I could be wrong if that's not

true across many other employer group based provider programs like that.

MS. JOUZA: And I'm not sure, and that may be, it may be a difference because if that's from a, you know, more of like that plan side too, we are focused on the specific provider side for this measure, or these measure versions. So that may impact kind of where or how that's documented.

But again, at some point, in order to report on this program, there is the acknowledgment and the acceptance by the eligible provider, eligible clinician to actually put through the actual coding in order to capture the numerator.

Similar for the EHR version utilizing the LOINC, SNOMED concepts, CPT if those are used for the actual interventions themselves.

MS. TIERNEY: Yes, I was just going to also add to what Jamie said. I mean, I think given that this measure's in the federal reporting programs and there's a penalty for not

reporting, I think there's an incentive for providers when they've done this to report that they've done this.

So I think that if it is done at the provider level, they probably would very likely report that they've done it. I understand what you're saying that it might be captured in other ways, but I imagine that they would plan, identify some sort of a work around so they could still get credit if you will on the measure because there is this penalty for not reporting.

MEMBER PINDOLIA: Right. And that's where I guess where I'm kind of in conflict with the gap results because it's only 22 percent you're saying are reporting. But then there's this penalty.

And the ones that report, there's really no gap. So that's where I'm trying to figure out there's no gap because we're missing data that's already captured in other ways, or there really is no gap.

CO-CHAIR PINCUS: Yes. Well, I think

one of the issues is, just to be clear, we're not looking at that gap based upon the program of reporting. We're looking at the gap in terms of this measure.

So that's, you know, and we may be looking under the lamp post with regard to this program. But it would be useful to have broader information about the overall on a population based level.

So Mike, Jeff, and David.

MEMBER TRANGLE: I want to speak to the gap for people who have mental health and CD services. I'm part of a collaborative in Minnesota that's with the Lung Association and DHS. And we mirrored something that I think was going on in North Carolina.

So there were two states that essentially over the last at least ten, maybe fifteen years have shown, look, with our data that the rates of folks smoking have gone down for every subgroup except people that have mental illness and CD problems where it's been totally

flat and there's been no progress.

And maybe you have some similar data that you were talking about. We, too, have been recommending that in the broad based measures that we do look at folks with mental illness and CD as a disparate group so that it could be tracked in a whole host of different measures, lifespan, other kinds of things, not just tobacco cessation which was one of the key drivers for lifespan.

But there is data saying there, for the group that we're attempting to look at, behavioral health patients, there is a real performance gap, but it gets lost in the shuffle if you're not subdividing out our population.

CO-CHAIR PINCUS: Jeff?

MEMBER SUSMAN: Briefly. When do we eliminate a measure? Do we need it to be 95 percent within, 99 percent? My sense is that this data isn't representative of the population as a whole and that the 21 percent or 22 percent that actually are reporting are perhaps different

than the other approximately 80 percent.

But without that data, we're sort of left in this well this feels like we should still be measuring it, there's certainly some disparities, some populations where we're not doing as well. At this rate, we'll never get rid of a measure.

CO-CHAIR PINCUS: David?

MEMBER PATING: Well, given that the USPTF just moved the measure up from number three to number two and three, I think it's -- moved down or something like that. But anyway, yes. It's, I think the need is growing greater.

I wanted to make sure that this measure, that we put in comments the issue of tobacco screening versus nicotine screening with the e-cigarette craze. I think we really need to look at whether this is the time and place to expand the definition to catch the next wave of the epidemic.

The second thing would be I would also just verify that I think public sector we still

have high rates of smoking. Now the veterans, I know you guys have high smoking in the vets, and mental health population.

So while health plans and primary care is probably doing well overall, there's the subsegments. And that's the way organizations should re-look at it. They should be segmenting their overall measures because as the difference in the total measure gets narrower, the separating yourself really depends on your margins.

So health plans are going to have to look at those anyway. But I think your comment is around the nicotine and e-cigarettes and expanding the definition.

CO-CHAIR PINCUS: Okay, so we're going to be dealing with the measure as it is, and obviously we can make recommendations going forward. But so Raquel and Mike, are there additional comments on gaps that have not been made before?

MEMBER PATING: I don't think it has

to do with gaps.

CO-CHAIR PINCUS: Okay. Connie.

MEMBER HORGAN: Well, I just want to add on the e-cigarette and the nicotine issue and make the point of the intersection with age and the phenomenal growth with adolescents so that if that does get considered in the future that that be taken into account.

MS. TIERNEY: Just to I think emphasize something you said earlier, I think there's different sources of data for the gap. You know, ideally you would have data from the measure, and we do have some data from the measure.

But I would argue it's not nationally representative because it's only 21.5 percent of eligible providers who could have reported on the measure.

We did also include in the submission form data from the medical literature around the provision of cessation interventions and the screening. And the rates there are not great,

you know, around 65 percent.

It's somewhat dated, you know, but it is around 65 percent were getting screened, and then about 30 percent were getting the intervention.

And additionally, you know, on top of that there's also the disparities that I think you know are additional --

CO-CHAIR PINCUS: No, I think your point is well made. And I think also you could add with, you know, moving from PQRS to MIPS, that's going to, intended to expand from the 22 percent to a much larger proportion. So that gives a view.

So why don't we move to vote with regard to gaps.

MS. QUINNONEZ: Voting is now open for performance gap of Measure 3225. Option number one, high. Option number two, moderate. Option number three, low. And option number four, insufficient.

Performance gap of Measure 3225.

Option one, high. Option two, moderate. Option three, low. And option four, insufficient. Thank you, all votes are in and voting is now closed.

For the performance gap of Measure 3225, 26 percent voted high, 6 individual votes. Sixty one percent voted moderate, fourteen individual votes. Nine percent voted for low, two individual votes. And four percent voted for insufficient, one individual vote.

For performance gap of Measure 3225, this passes the measure criteria.

CO-CHAIR PINCUS: Okay, so let's move on to reliability. Do one of the members of the group that were reviewers want to comment on reliability? Anyone? I'll have to call on somebody. Mike?

MEMBER LARDIERI: I'll go ahead, but it seems reliable. Reliability is not my specialty, Harold, I think you know that. So --

But from what I'm reading here, yes it does seem reliable. It's been around for a

while. So I think it's safe to say that it's still reliable.

CO-CHAIR PINCUS: Okay, Dodi?

MEMBER KELLEHER: And I think again this is a maintenance measure where they did do, show the 2015 and sampling as well as the testing for 2017 evaluation. So, and in I think it -- to echo, it seems reliable. Meets the algorithm, and so I think that's pretty much it. And a nice big sample.

CO-CHAIR PINCUS: Any further comments about reliability? Okay, so why don't we vote on reliability?

MS. QUINNONEZ: Voting is now open for the reliability of Measure 3225. Option number one, high. Option number two, moderate. Option number three, low. And option number four, insufficient.

We are voting on the reliability of Measure 3225. Thank you, all votes are in and voting is now closed.

For the reliability of Measure 3225,

48 percent voted high, that's 11 individual votes. Fifty two percent voted moderate, twelve individual votes. Zero percent voted for low, and zero percent voted for insufficient.

For the Measure 3225, reliability, this passes the measure's criteria.

CO-CHAIR PINCUS: Okay, let's move on to validity. Comments from the reviewers with regard to validity? Dodi?

MEMBER KELLEHER: So there was the 2017 evaluation was again face validity. And though I don't think anything I have to say makes it not valid or a lot less valid, I was concerned about the ten and the more importantly that there seemed to be more disagreement about the level of agreement. And I wondered if you wanted to make any comment about that.

MR. SEGOVIA: Yes, I can speak to that. So it's unfortunate we only were able to get ten responses since the submission, but we actually have added additional responses. So we actually got responses from 29 folks since the

submission and raised the validity score to 76 percent, strongly agree to agree.

We can also strongly agree as well as agree that the measure was valid.

(Off-microphone comments.)

MR. SEGOVIA: I'm sorry? 76 percent. Yes, and I would be happy to add that to the forms.

MEMBER KELLEHER: I have one more comment. The other comment I want to make is about threats to validity. And my, I sort of have a question or an issue about what would constitute a reason, a medical reason not to screen, even if one's life expectancy is short, one was terminal, one had heart disease and likely a short lifespan.

Why would you not want to screen and offer intervention if the screening were positive for someone who, you know, was certainly not going to get better being allowed to smoke, or at least being offered the opportunity.

I say this because in my own

experience I've seen, because it's an addiction, I've seen people who were outside hospitals having their last smoke before they got, you know, angioplasty.

So I was just curious as to why that exclusion. I don't think it's a big threat, but I don't see why.

MS. TIERNEY: Yes, that's a good question. I do think the greatest example that we discussed with our expert workgroup and more recently our technical expert panel, it really does relate to the limited life expectancy issue.

I think there's, you know, a comfort in knowing that if there's a valid reason that as a physician you felt that you didn't want to screen someone, that you could do that.

And so I think that's the largest reason it's there. And I think that we heard from, you know, the palliative and hospice medicine community that this is an appropriate exception for us to have.

I do think that the rates of the

exceptions as we've, you know, from our analysis are still pretty low so that they're I think across the different modalities they're all less than one percent.

So I don't think that people are using it inappropriately, but I do think it gives the option, and our expert workgroups felt that it was, you know, more comfortable with it being there if there was a good reason that you chose not to screen someone.

MEMBER PATING: So I just have a question regarding 2B5, meaningful difference, and then also comparing this to the chart above that under reliability.

So under reliability, you show that you're able to show that 25 percent of claims providers did not meet threshold. I guess meaning that they didn't either screen or they didn't provide treatment.

And to me that seems like significant difference and that gap that I would want to close. And then down here on your meaningful

difference, however, your 25th percentile on claims is .97 and your 75th is 1.

So at one level you're looking at individual providers, and another level you're looking at claims. I guess how do you recommend this measure be used and which meaningful differences are more meaningful.

MS. TIERNEY: So can you point me to what you're talking about in regards to the --

MEMBER PATING: So 2B5.

MS. TIERNEY: 2B5.

MEMBER PATING: 2B5 under validity which is threats to validity you see claims data with a standard deviation of .1, and the 25th percentile is 97 percent. And the 75th percentile is 100, right, or 1.0.

And then if you go up to the reliability section, you'll see claims where you had 190 events, 53,000 providers. And you said 25 percent of the providers didn't meet thresholds which means they didn't do something.

MS. TIERNEY: Right, so the threshold

there, just thank you for showing me where it was. I'm sorry, I think I'm looking at the wrong document. But the threshold there refers to the threshold that was determined to conduct the signal to noise analysis.

So anyone with less than ten events was excluded. So 25 percent of providers in claims did not have in the claims sample that we provided did not have ten events.

The measure is intended to be used at the individual clinician level. And so all the data, even though it may be the source of the data is claims, it's all physician level data.

MEMBER PATING: Okay.

MS. TIERNEY: Does that help a little bit?

MEMBER PATING: Well then you still have the meaningful difference problem which is can you really separate low performers from high performers either at the individual or the plan level. And it just, I think it's something that needs to be addressed, whether retired or

expanded to a broader definition set.

MS. TIERNEY: So I think the challenge with the meaningful differences in performance with our data sample here is that the mean performance rate especially in claims was very high because as we've discussed there were, you know, not that many physicians reporting on the measure, or not as much as we would like.

So I think that's why the meaningful differences are less, I guess, noteworthy within the claims sample because of the high performance rate. I think if we had something that was, you know, more representative, and I think in the sample above with the registry where the performance rate is lower, the meaningful difference is more significant.

And I think similarly, you would see the same thing in the EHR data we have because of the data from that, the mean performance rate is even lower. So I think it's a function of the data that we received and the fact that the claims performers, the limited sample of people

who were reporting in claims, are reporting very well.

CO-CHAIR PINCUS: Again, I think one of the issues is that it's unfortunate that the data's so limited to the groups that have agreed to participate in a program which means that they probably chose this because they do well on it.

And so that's part of the issue. Mady, you still have a question?

I had one other question about validity which goes back to the origination of this which is how do we know that the, for those who are identified as needing cessation interventions are actually getting evidence based interventions?

MS. TIERNEY: SO that's a great question. The eMeasure requires either pharmacotherapy, and we define that as, well our eMeasure defines that but our claims registry version doesn't define that. It's through attestation which Jamie can speak to.

The way we define counseling is brief

counseling, three minutes or less. Of course someone could do more.

I realize from the evidence that, you know, the greater the intensity of the counseling, the greater impact on tobacco cessation. So ideally yes, we would want them to do more.

But given the focus of the measure, you know, its national impact in trying to be something that could be easily done within primary care.

You know, we elected to define counseling as three months or less because that does have an impact based on the data from the guidelines, but albeit not as significant as if it was longer or if it was in several different instances, you know, multiple times.

CO-CHAIR PINCUS: In an ideal world, you would want this measure to number one, be validated by even in this 22 percent of physicians participating in the program to see whether or not for those patients, even those

clinicians, whether there was actually a higher rate of smoking cessation.

And in fact, actually, you probably would want to move toward a measure that actually includes actual smoking cessation.

But, and that may be aspirational at this point, but it's something that, you know, at least my point of view, should this measure come up again for maintenance, that that would be an expectation.

MS. TIERNEY: I definitely think your point's well taken. I think that's something we would certainly want to move to and certainly a topic that we can discuss with our TEP as we review the measure annually and just determine sort of future directions with it.

CO-CHAIR PINCUS: Any discussion about validity? Okay, let's --

MEMBER MARK: I had one point. The US Preventative Services Task Force gives smoking cessation an A rating, right? So isn't that enough validity to suggest that smoking --

CO-CHAIR PINCUS: I couldn't hear the last part --

MEMBER MARK: I was saying that it seems like enough validity to suggest that smoking advice works if the US Preventative Task Force reviewed all the materials --

(Simultaneous speaking.)

CO-CHAIR PINCUS: Well, we're talking about the validity of this measure, not just --

MEMBER MARK: I know, but this measure is basically just measuring whether the provider gave advice to someone who is smoking. And the US Preventative Services Task Force says that giving advice is highly effective.

So it seems that it's a direct, you know, it's tautological this measure can't be sound even then. But maybe I'm missing something.

CO-CHAIR PINCUS: No, I'm saying but the measure itself says evidence based. That's in the description.

MEMBER MARK: Well, I guess the other

point is, you know, I think it's interesting that this measure is, if I look at the claims data, I'm kind of surprised how high the rate of compliance is because if you look at the screening and brief intervention for alcohol use which is also a claims based measure, it's teeny.

And so I'm wondering what the difference is. And maybe the difference is because you only have to provide three minutes for this measure, and for the SPI measure it has to be at least 15 minutes.

So I wonder if, I mean, I think that it would be interesting to look at whether it has to be effective, if it can be effective at three minutes. You know, but I don't think that actually negates this measure, I just think it would be an important way to decide.

CO-CHAIR PINCUS: Mady?

MEMBER CHALK: This is just a brief comment that takes us a little backwards. But the US Preventative Services Task Force definition of not only the smoking cessation

interventions that it got an A rating, but also that the question about the exclusions earlier.

All of it rests on what the USPTF has said. The issues that arise about how can a measure be constructed such that -- I lost my train of thought. Skip it, I'll go back to it later. I lost my train of thought.

CO-CHAIR PINCUS: Let's move to voting on, oh Tammy, you still? Let's move on voting on validity.

MS. QUINNONEZ: Voting is now open for validity of Measure 3225. Option one, moderate. Option two, low. Option three, insufficient.

For the validity of Measure 3225, Option one, moderate. Option two, low. Option three, insufficient.

DR. LUSTIG: And again, just a reminder, the highest rating is moderate because the validity is based on face validity testing.

MS. QUINNONEZ: Looking for one more vote. Thank you, all votes are in and voting is now closed. For the validity of Measure 3225, 87

percent voted moderate, that's 20 individual votes. Nine percent voted low, that's two individual votes. And four percent voted insufficient, one individual vote.

So for the validity of Measure 3225, it passes the criteria.

CO-CHAIR PINCUS: So moving right now moving to feasibility. Any comments on feasibility? Anita?

MEMBER PINDOLIA: Just one quick comment. For, and I was trying to quickly identify this, the numerator, you get counted if you screened and then the intervention could be either the three minute discussion or use of a smoking cessation pharmacotherapy product? Does that include OTC nicotine products?

MS. JOUZA: I'm trying to recall from our RxNorm for the eMeasures. I do believe that it does because we do offer kind of like a plethora. There's a patch, there's a pill, et cetera, etc.

MEMBER PINDOLIA: So then the second

question for the feasibility, how does that get captured, because we don't capture that in Rx claims.

MS. JOUZA: How does it get captured in the claims registry? Again, unfortunately it is that attestation for it. It's going to be documented somewhere in that patient's record.

MEMBER PINDOLIA: Oh, just that the doctor said --

CO-CHAIR PINCUS: An attestation.

MEMBER PINDOLIA: Okay, thanks.

MS. JOUZA: Exactly. And it could be --

(Off microphone comments.)

MS. JOUZA: Yes. And it could be either/or or a combination of both.

CO-CHAIR PINCUS: Any other comments about feasibility? Okay, ready to vote on feasibility.

MS. QUINNONEZ: Voting is now open for the feasibility of Measure 3225. Option one, high. Option two, moderate. Option three, low.

And option four, insufficient.

Looking for one more vote. Thank you, all votes are in and voting is now closed. For the feasibility of Measure 3225, 52 percent voted high, 12 individual votes. Forty eight percent voted moderate, eleven individual votes. Zero percent voted for low, and zero percent voted for insufficient.

This passes the criteria of feasibility for Measure 3225.

CO-CHAIR PINCUS: Okay, now use and usability. Any comments on use and usability? Mike?

MEMBER LARDIERI: Yes, I'm not sure if it comes up here or not, but I think it would be more useful if we were able to put another question at the end that gets to the smoking on the panel question which is one of the other measures.

So if you would identify the percentage of the patients who are actually no longer smoking as well as those other three

questions, I think that would, me as a provider, make that much more useful because then I know what percentage of my population is smokers and not smokers.

And if I've done anything, I can just look and see how I've brought that number down.

MEMBER KELLEHER: And just to pound the nail in all the way, I think all the tobacco related measures need to think about changing their titles and including nicotine delivery rather than tobacco to incorporate, keep up with the times here, especially as it relates also down to the pediatric level.

CO-CHAIR PINCUS: So if the staff could keep a record of, like, all these comments that we're making in terms of future improvements to this and to the other tobacco measures, nicotine measures.

Any other comments about usability and use?

(Off microphone comments.)

CO-CHAIR BRISS: So on the last point,

on the last point you think about the pros and cons of that. So one of the reasons that a lot of measures get specified for just inhaled tobacco use is first that inhaled tobacco use is the public health most important one.

And if you want to measure all the other tobacco things, you greatly increase the measurement burden because then people have to ask about hookahs and there's a whole long list of other things that you might make people ask about.

And so think about the trade-offs of what you ask for. At this point, I'm not voting. My tobacco people at CDC would prefer to have the whole long list, truth is. But there are a lot of healthcare systems that don't agree with that.

CO-CHAIR PINCUS: So maybe our recommendation is figuring out how to solve that problem. Any other comments about usability and use? Okay, let's vote.

MS. QUINNONEZ: Voting is now open for usability and use of Measure 3225. Option number

one, high. Option number two, moderate. Option number three, low. And option number four, insufficient information.

Thank you, all votes are in and voting is now closed. For the usability and use of Measure 3225, 50 percent voted for high, which is 12 individual votes. Forty two percent voted moderate, that is ten individual votes. Eight percent voted low which is two individual votes. And zero percent for insufficient information.

For usability and use of Measure 3225, this passes the measure criteria.

CO-CHAIR PINCUS: So are there, do we have to deal with any kind of competing?

DR. LUSTIG: There's nothing competing. And so as we said before, we're going to talk about this in the scope of our overall portfolio tomorrow.

CO-CHAIR PINCUS: Okay. So let's vote on endorsement, but before we do so, are there any final comments with regard to endorsement issues? Mike?

MEMBER TRANGLE: Question as much as a comment or recommendation or wishful thinking, I don't know what the right title is. But could we either pass it with a recommendation or in suspense that's my wishful thinking, that it is analyzed so that we can look at the results with the BH subset and look for improvements so it doesn't remain lost in the shuffle and we wonder whether there's even a gap or that kind of thing.

CO-CHAIR PINCUS: So when did this come up for renewal of their maintenance?

DR. BURSTIN: I will point out, and Harold knows as well, and NCQA will be up shortly that there are a whole set of CD risk production measures specifically for the behavioral health population.

There's no reason you couldn't suggest to the developer that they add, that they begin to explore adding the strata. It doesn't have to wait three years, that could be part of our annual review process that we recognize the importance of this and we would like to see them

begin to include a stratum for serious mental illness.

CO-CHAIR PINCUS: Yes, in a sense creating expectations about other measures that are looking at this kind of thing, you know, these phenomena that they sort of move to the sort of next level of expectation in terms of being able to measure things that are more meaningful.

MEMBER TRANGLE: It's really not just a intellectual thing because what we're finding when we're trying to do the work is that the general approaches that have good evidence that they've improved things for the general population don't work for the behavioral health population.

And if it gets shown for the general population, then there will be more funding and more recognition that something different needs to happen for our sub-population.

It drives money, it drives attention, and potentially resources. So it's really not

just an esoteric kind of thing.

CO-CHAIR PINCUS: Okay, any further comments before we vote on endorsement? Okay, why don't we vote.

MS. QUINNONEZ: Voting is now open for the overall suitability for endorsement of Measure 3225. Option number one, yes. Option number two, no. Thank you, all votes are in and voting is now closed.

For the overall suitability for endorsement of Measure 3225, 100 percent voted yes, that's 24 votes.

CO-CHAIR PINCUS: Okay, so why don't we move on to the next one which is the eMeasure version of this. Okay, take it away.

MS. TIERNEY: You know, I don't think we have anything additional to add. In my earlier discussion we sort of described the overall importance of the measure and --

CO-CHAIR PINCUS: Right, but if you say something about how, what you've done to specifically adapt this for the measure.

MS. TIERNEY: All right, Jamie, would you speak to that?

MS. JOUZA: Sure. So this measure is kind of broad in nature in that we don't have those targeted populations as we've been discussing. And so our denominator essentially captures all of those patients who are 18 years or older.

In terms of that numerator though, in order to adapt it for the eMeasure, we have had to codify and use some of those recommended clinical vocabulary LOINC to capture our screening tool, or screening questions. RxNorm to capture the pharmacotherapy agents as well as CBT and SNOMED to capture the actual procedure of the cessation intervention.

And so in order to kind of take it from that attestation piece with the CPT2 code as has been included in the PQRS program since the measure was introduced, which we just said was back in, like, 2009, in some form.

And so that has been kind of expanded

and able to kind of capture and really look at what is being required for the measure. And so there's no longer that attestation you have to show that you've done, you've screened, you have a tobacco user status, tobacco non-user status.

And so that's been adapted and used as different coding terminologies.

CO-CHAIR PINCUS: Just, I might have missed this but how do you capture OTC?

MS. JOUZA: Those are going to be through RxNorm codes. And I think there would be a prescription. So we allow for a medication order or a medication active.

And so if the physician were to give, you know, they recommend that you go pick this up, then they could do that. That would still be documented kind of in that.

CO-CHAIR PINCUS: So that would be documented in the --

MS. JOUZA: Yes, we do allow, like I said, if a patient is actively taking the cessation pharmacotherapy or is ordered to do so

at that particular screening, that counts as the intervention.

CO-CHAIR PINCUS: Any questions on the clarification here? So is there any objection to having further discussion with regard to evidence? Any? Okay, so we can move on beyond evidence to gaps. And here the gaps are become relevant to the particular groups that are being evaluated. And is there data about that?

MS. TIERNEY: We do have data from the sample that we received from CMS that we conducted our testing analysis on that, chose lower rates of performance for this measure, come here to the claims and registry versions of the measure.

CO-CHAIR PINCUS: What there was lower rates?

MS. TIERNEY: Lower rates, yes.

CO-CHAIR PINCUS: What magnitude?

MS. TIERNEY: I don't have it right at my fingertips.

(Off microphone comments.)

MS. TIERNEY: Thank you. 76.83.

CO-CHAIR PINCUS: Any comments about gaps on this measure from any of the reviewers?

MEMBER HORGAN: I think the only issue is that it performs slightly less well.

CO-CHAIR PINCUS: What was that?

MEMBER HORGAN: I think the only gap is that performance is less, and we just listened to those numbers. But everything else seemed similar to the extent of the study.

CO-CHAIR PINCUS: Okay.

MEMBER MAZON JEFFERS: Will it attribute that lower performance to the fact that you've eliminated the attestation aspect of the measure, and you are looking more at verifiable evidence?

MS. JOUZA: Yes, I think, and also kind of the adoption of the EHR in moving to that comfortable, you know, claims registry to kind of requiring, again these are different coding terminologies than many are comfortable with or used to.

And so that could also contribute to I think the lower performance. While they understand, you know, how to report and which codes are applicable to the measure.

But again, because you can capture and the way that the measure is structured, the eMeasure is structured, we have our screening and non-user, and then we have our screening and user and intervention.

And so with that, CPT2 in the claims registry version 3225, that kind of lumps it all together. And so now we are actually making that computer requirement that you have to have this and this and this.

CO-CHAIR PINCUS: So just a question. So I think, I can understand why the rationale why it might be a lower measure. But does that also suggest that during this phase it's also created unreliability if part of the problem is sort of the, you know, the application of the methodology for capturing the data?

MS. JOUZA: I think that there are

oftentimes questions that we received that implementers, vendors themselves are kind of unsure how to kind of pull things together and want to really ensure that they are meeting their requirements of the measure.

So there might be a misunderstanding and potentially lower reliability. But again, this measure have been a part of, and there are other tobacco related measures that are included in the program, including the smoking objective that was part of meaningful use.

And so I think that there is better appreciation for what those requirements are, again for EHR specific reporting too.

CO-CHAIR PINCUS: Okay. Any further discussion about gaps?

MEMBER MAZON JEFFERS: I'm sorry, I remembered my question. Is this for use only in primary care settings? I'm sorry, I didn't do the in depth review on this one.

MS. JOUZA: It is not. This measure actually includes a variety of settings, a

variety of encounter, different types of encounters. And so we do have the office primary care physician settings. We do have ophthalmological visits, occupational therapy, speech hearing evaluation encounters.

And that was all based on, as Sam had mentioned before, when we developed the measure initially, we had a multi-disciplinary workgroup. And so we had a lot of those specialties kind of at the table and subsequently received some requests from CMS and we received some requests from specialties in particular who requested that they have their encounters added.

MEMBER MAZON JEFFERS: But it doesn't include behavioral health settings?

MS. JOUZA: It does.

MEMBER MAZON JEFFERS: It does?

MS. JOUZA: Yes.

CO-CHAIR PINCUS: To vote on the gap issue.

MS. QUINNONEZ: Voting is now open for Measure 3185, preventative care and screening

tobacco use screening and cessation intervention eMeasure. We're voting on performance gap.

Option number one is high, option number two is moderate, option number three is low, and option number four, insufficient. For the performance gap of Measure 3185, option one high, option two moderate, option three low, and option four insufficient.

Thank you, all votes are in and voting is now closed. For the performance Measure 3185, performance gap, 33 percent voted high which is eight individual votes, 67 percent voted moderately which is 16 individual votes, zero percent voted for low and zero percent voted for insufficient.

So performance gap passes the criteria for this measure.

CO-CHAIR PINCUS: Okay, let's move on to reliability. Any of the reviewers who reviewed it want to comment on reliability? Connie?

MEMBER HORGAN: The reliability scores

were similar to the previous measure which makes this very easy to do. The reliability at the minimum number of events was 181, and it was .99 at the average number.

CO-CHAIR PINCUS: Any additional comments with regard to reliability? Mike?

MEMBER LARDIERI: Yes, I just have one. Where does dental fit in? Does that fit in under clinician, office clinic? I missed that on the other one too. It doesn't look like dental is involved.

MS. JOUZA: We don't specifically address the dental community within the measure itself.

MEMBER LARDIERI: Is there a reason why not or are you going to include that? I mean, dentist is a great place to do that.

MS. TIERNEY: Yes, I mean, I think it's something we could consider in future updates. As Jamie said, we've had, you know, we had this broad workgroup. It did not include dentists when we developed the measure. So some

of the people who were on that workgroup certainly felt that, you know, it was appropriate for their setting of care. So they, you know, encouraged us to have that setting included within the measure specifications.

We have received subsequent requests from other professionals that have also allowed those codes to be added once we've added those through a process that we follow. But we've never had a request from the dental community. But I think it's something we could certainly pursue.

You know, of course the appropriateness of it with our technical expert panel and then also involving the dentists, the dental community to see if this is something that they would want to report on.

DR. BURSTIN: I'll just mention NQF did a report several years ago specifically on the future for dental measurement and this whole area of screening for oral cancers is so prominent. Anyone who's been to the dentist at

8:00 a.m. this morning, it was a good full five minutes of my exam. It seems like such a logical place to put smoking as well. So I would be happy to share that report with you guys.

CO-CHAIR BRISS: So a couple of comments on the specs that I would like to see improved over time, actually doesn't reflect, doesn't deal with the stuff for today.

But I would like, I would love to see you move toward a less restrictive denominator. So the at least two visits, the people who are seeing you pretty routinely aren't the people that I'm most worried about actually.

So I would like to see the denominator be less restrictive than that in terms of how many visits are required. I would also like to see as tobacco prevalence in the population goes down, your performance on this measure is more and more about screening people for tobacco use who aren't smokers.

So hypothetically, if you were to encounter a population that had a national

average tobacco use that was 15 percent, you did a perfect job screening and a zero job of actual treating the actual smokers, you could score 85 percent on this measure as it's currently specified.

And so the gap is increasingly not on the screening side, it's on the treatment side. And I would much like to see you separate the screening from the treatment and give us separate scores for the two pieces.

MS. JOUZA: So that's a great point, and just to jump in, that is actually something that we are kind of modeled and put forward to consideration for, by CMS.

And so they're actually reviewing that proposal to kind of split it out so that we can actually see those rates for the actual screening, and then of the screening, of those who were screened and were tobacco users in our denominator, who got the intervention so that we can get that and actually see overall where the gap is.

CO-CHAIR PINCUS: I think, just a question for Helen. You know, I think that makes a lot of sense. But it's important to at least in my mind that it be combined as a single measure but with two different so that it's not like people can pull measures out of the arm room and carry them and say oh, we're just going to do screening.

MS. JOUZA: And that's exactly what we're trying to do.

CO-CHAIR PINCUS: Any other comments about reliability? Okay, we're ready to vote on reliability.

MS. QUINNONEZ: Voting is now open for the reliability of Measure 3185. Option one, high. Option two, moderate. Option three, low. And option four, insufficient.

Thank you, all votes are in and voting is now closed. For the reliability of Measure 3185, 21 percent voted high, 5 individual votes. Seventy nine percent voted moderate, nineteen individual votes. Zero percent voted for low,

and zero percent voted insufficient.

Reliability of Measure 3185 passes this criteria.

CO-CHAIR PINCUS: Let's move on to validity. Comments on validity? Dodi?

MEMBER KELLEHER: I just wanted to note that the Bonnie testing was 100 percent.

CO-CHAIR PINCUS: Connie?

MEMBER HORGAN: And I was thrilled to find out that the name Bonnie meant nothing.

My comment relates to something that we've discussed before on the previous measure that's allowing medical reasons. And the comment that was given is that there have been concerns expressed about this being gamed.

The explanation that was given is that it's so low it's not worth worrying about. I just want to point out that as we move forward, and we're only dealing with a relatively small proportion of physicians who aren't using this measure as it becomes more widely used, I don't know that you can assume that it's immune from

gaming.

And that is something that I think we should continue to look at because it is extremely vague in terms of the, right?

CO-CHAIR PINCUS: Is there another medical reason other than limited life expectancy that anybody had in mind?

MS. TIERNEY: We have had this discussion with our expert panel just recently in fact. And there was general agreement that there are some situations that you may not find it appropriate at that visit to perform the screening.

We didn't get specific examples to add to the measure. We asked are there, you know, some that happen frequently enough that we could add to the measure just because we do get asked this question, are there other examples.

We didn't get specific examples, but the general consensus of the test was that there are situations and, you know, maybe somewhat rare but there are situations apart from limited life

expectancy where you might not want to screen.

And I think it was more focused on at that visit maybe because at that visit you would be focused on other things.

CO-CHAIR PINCUS: Connie?

MEMBER HORGAN: Just following up on that, and I am not a clinician so I defer to the clinicians in the room. But I have heard that argument used specifically for behavioral health and that you would not want to screen, or mental health and other addictions because of certain situations. So I just want to loop that back.

CO-CHAIR PINCUS: Peter?

DR. SAXON: Hi, this is Andrew Saxon, I'm with the American Psychiatric Association and work with the PCPI. And so I just want to very strongly argue against that point of view.

Most of the remaining tobacco users are mental health patients. That is the place where we need to be screening the most. And there really is no reason why that cannot be included in a mental health visit.

Even if a person is suicidal, tobacco, smoking doubles the risk for suicidal behavior, so it should be inquired about as a potential risk factor for suicide.

MEMBER HORGAN: I would like to clarify that I'm agreeing with you. The only point was that if it's left vague as a medical condition, it opens it up to gaming. So one should protect against that happening.

DR. SAXON: I'm sorry if I misunderstood your comment.

MEMBER HORGAN: Important to clarify it.

CO-CHAIR BRISS: I just wanted to endorse the idea of less exclusions. Generally, I think less exclusions are better. And if that means that we're not always going to score 100 percent all of the time, I'm fine with that. Especially if the exceptions are rare.

CO-CHAIR PINCUS: Yes, it's only .4 percent. Okay, I think then we're ready to vote on validity.

MS. QUINNONEZ: Voting is now open for validity of Measure 3185. Option one, moderate. Option two, low. Option three, insufficient. All votes are in, voting is now closed.

For validity of Measure 3185, 79 percent voted moderate, 19 individual votes. Twenty one percent voted low, five individual votes. And zero percent voted for insufficient.

For measure 3185, this measure passes the validity criteria.

CO-CHAIR PINCUS: Okay, let's move on to feasibility. Any comments about feasibility? Connie?

MEMBER HORGAN: I would like to ask if you could offer more explanation about the testing you did in two different systems and then you said in one system, the data elements, only 17 of the 26 elements were currently feasible. And then it was a statement about they were judged to be feasible in the next three to five years. Could you expand on that?

MR. SEGOVIA: Sure, I can expand on

that. So in terms of the feasibility in some EHRs, we noted that some providers don't use particular encounter counts.

So if your internal medicine provider, you may not be able to use behavioral health codes. So that contributed to some of the not being able to use some of the encounter codes for that.

Also I want to note that some of the exclusions were unable to be captured in structured fields. So most folks use free text to document those.

MEMBER HORGAN: I was more commenting on how do you know it's going to be better in three to five years? It's just, like, a random number, everything gets better in three to five years.

MR. SEGOVIA: Which system was --

(Off microphone comments.)

MR. SEGOVIA: It was what?

PARTICIPANT: It was an Epic --

MR. SEGOVIA: Implementation?

CO-CHAIR PINCUS: Epic was -- so the VA was able to --

MR. SEGOVIA: The VA was, okay.

CO-CHAIR PINCUS: Other comments with regard to feasibility? I guess, you know, I don't know. Mike, you would probably know in terms of market penetration. But, you know, Epic I think is the largest. And it would take them three to five years, does this have any implications for other vendors?

MEMBER LARDIERI: Some other vendors can probably do it faster. Epic has a certain way of implementing changes. They come out, like, once every six months, and they do the whole world of Epic at once, and then they move on.

And if you're in the queue at that point, it comes on. If you're not, you have to wait again. So a lot of it has to do with that. But other vendors could do it sooner. I think three to five years, this could happen sooner than three to five years if organizations wanted

it.

Again, it gets to the issue of there's another issue. I mean, do the vendors have to pay in order to make this happen, and then those costs then go down to the provider. And maybe I don't want to pay for that. So you know, that also is implied with these things.

CO-CHAIR PINCUS: Any other comments about feasibility? Okay, I think we're ready to vote.

MS. QUINNONEZ: Voting is now open for the feasibility of Measure 3185. Option one, high. Option two, moderate. Option three, low. Option four, insufficient.

Looking for two more votes. Thank you, all votes are in, voting is now closed. For the feasibility of Measure 3185, 29 percent voted high to 7 individual votes, 71 percent voted moderate to 17 individual votes, zero percent voted for low and zero percent voted for insufficient.

For the feasibility of Measure 3185,

it passes this criteria.

CO-CHAIR PINCUS: Okay. Let's move on to use and usability. Any comments about use and usability? Connie?

MEMBER HORGAN: I know Peter spoke about everyone at the CDC wants everything in this measure, but I think the fact that the implementers of this measure were asking for more information and addressing the issue of electronic, a continuing delivery system should be taken, given a high priority of this particular mechanism that's not being included.

So as we get into the future I think it might distinguish it more from some of the others.

CO-CHAIR PINCUS: Okay, any other comments with regards to usability? David?

MEMBER PATING: Since Peter got to make a speech I'm going to make a little bit as well.

So as pain as the fifth vital sign falls out, I would be hopeful that tobacco can

move in. I think tobacco should be every visit, every time, and it's like the suicide question for mental health and it's such a major public issue.

So with regards to that I think the issue because measures, we're showing really high screening rates, but as moving into treatment as Peter is mentioning is a really good goal, but absolutely realizing this is a relapse in disease we may even need to look at six- and two-year outcomes.

There's Doug Ziedonis and several national speakers that are moving towards the idea of tobacco registries that we really need to be thinking longer term with tobacco cessation, probably more relapse than not, and everyone knows it takes seven times to quit smoking or before you're effective.

So as you look to this development of this measure, broadening it, including more things, thinking farther out, it's still our number one killer. So thanks.

CO-CHAIR PINCUS: Thank you. Other comments on usability and use?

So I think we're ready to vote on usability and use.

MS. QUINNONEZ: Voting is now open for usability and use of Measure 3185. Option 1 high, option 2 moderate, option 3 low and option 4 insufficient information.

Thank you, all votes are in and voting is now closed. For the usability and use of Measure 3185, 42 percent voted high, ten individual votes, 58 percent voted for moderate, 14 individual votes, zero percent voted for low, and zero percent voted for insufficient information.

So for usability and use of Measure 3185, this passes the measure criteria.

CO-CHAIR PINCUS: So before we get to the voting of endorsement any final comments with regard to endorsement? Mike?

MEMBER TRANGLE: In my attempt to win the broken record award could we make sure we

stratify this?

CO-CHAIR PINCUS: I think there may be a lot of candidates for that so --

MEMBER TRANGLE: But I would just want to make sure that we really see how well this sub-population of behavioral health patients fare as this plays out over time. Just like I was saying it for the non-electronic version.

CO-CHAIR PINCUS: Thank you.

Other comments before we vote on endorsement? Okay, so let's proceed to vote on endorsement.

MS. QUINNONEZ: We are now voting for the overall suitability for endorsement of Measure 3185. Option 1 yes, option 2 no.

Looking for -- thank you. All votes are in, voting is now closed. For the overall suitability for endorsement of Measure 3185, 100 percent voted yes.

CO-CHAIR PINCUS: Nice work. We've gone through a lot of these in a row. So now we're going to consider Measure 3229. And I

guess, Peter, you're recused?

CO-CHAIR BRISS: I'm re-recused.

CO-CHAIR PINCUS: Re-recused.

DR. LUSTIG: My apologies about the earlier one, but Dr. Briss is definitely recused from this measure.

CO-CHAIR PINCUS: Okay. So if the measure developers can introduce themselves and give us a summary?

DR. MCNAIR: Good afternoon, everyone. My name is Tiffany McNair and I'm a division director at the CMS Innovation Center's Prevention and Population Health Group.

I'm joined by my colleague and the lead of our measure's portfolio, Dr. Katherine Sapra. We're very excited to be here today and really appreciate the opportunity to share our measure with NQF and also with the committee.

I'd like to start out just by saying that as a practicing OB/GYN myself, I really am very much encouraged by the promise offered by population based outcome measures such as the one

that we are seeking endorsement from you all today.

And we recognize that while our measure is relatively simple, the constant undergirding it is not. As we shift from volume to value and increasingly focus on enhancing quality within our health care system an emphasis on outcomes is logical and yet at times it seems as if moving in this direction is a Herculean feat.

And what we're offering today for consideration is really, as one of the committee members put in their preliminary analysis, a first step. And I think as many of you know, a journey of a thousand miles begins with that first step.

And we hope that we can provide clarity for you all in the technical merit of the measure and also have a fruitful discussion on its promise, not in isolation but in harmony with other related process measures such as the one that we just talked about.

We hope that this will help to advance quality population health, and of course a system in which disease is not only prevented but really health is promoted.

So the evidence for our measure is well supported that smoking and cessation activities are clear. We thank our predecessors for the prior measure for really articulating that well and really helping to clarify the ongoing performance gap.

What I will say in addition to that is that beyond the kind of in-office counseling and the provision on behavioral and pharmacol therapies as well as the linkages that are created to community based services, we also know that the evidence supports really increased care coordination and case management to strengthen that.

So for instance, a patient that's seen in your office ensuring that they have not only referral but strong connection and linkage to quit lines, for example, with additional phone

calls and follow-up not only from the outpatient setting but even upon discharge. We know that this increases uptake of successful interventions and therefore potentiates the patient's ability to stop smoking.

And yet despite all of these available tools as well as the fact that there are existing process measures, we know that smoking prevalence continues to persist and is a really widely variable across practices, and as many of you have mentioned already is higher among certain subgroups as well and we're not necessarily capturing that in the current milieu.

So given the strong evidence base and the performance gap, we're proposing what is relatively simple but a far-reaching measure to capture the intermediate outcome of smoking prevalence. We just want to provide a few clarifications for you to our submission package and then we're looking forward to the discussion.

3229 is collected via electronic health records and it captures the percent of

adult patients, specifically those that are age 18 or older, who are current smokers among those patients who are screened for smoking.

Now the timing for collection is in line with the existing measure NQF 0028 which is now 3225. The measurement period is one year, so basically a patient must have had as you all know a qualifying encounter within that measurement year. However, their smoking status could have been ascertained within the 24 month prior, so either in the measurement year or 12 months before that.

In terms of exclusions, as with NQF 0028 one could have been excluded from this if for medical reasons if they weren't screened, such as limited life expectancy as we also heard. But in our case you could have also been excluded simply because a provider did not employ universal screening, and we'll talk some more about that.

Regarding the numerator and denominator, all patients within the denominator

must have been screened, right, and all patients in the numerator must have had a positive screen. This is an important distinction from 0028, so with the denominator in 0028 it includes the entire patient panel. Ours just includes those patients who were screened and therefore you know that they're either a smoker or a nonsmoker. That's the denominator.

In the numerator with our measure we stop at whether or not that person is a smoker. So you screen positively, you're in the numerator. With NQF it goes on to further query as to whether or not that person received counseling or interventions.

Now the data for the proposed measure also was obtained via the Medicare PQRS reporting mechanism and a sample was drawn from over 70,000 patients from nearly 400,000 eligible providers and across all regions of the U.S.

One point that may not have been fully articulated in our submission was that testing was conducting on eligible providers with at

least ten patients and among those who reported smoking status on at least 50 percent of their patients.

Findings from the testing demonstrate that our measure is highly reliable -- we hope that you are in agreement -- regardless of region or practice geography and regardless of patient panel size.

Specifically among those eligible providers who did see at least ten patients, the mean reliability was 0.89 and was even higher among those with at least 100 patients.

Another piece of late-breaking news that we'd like to provide for you on face validity because we were also encouraged on this piece was that our technical expert panels did weigh in on whether or not this would be a good measure of quality, so to distinguish between poor and good quality of health care.

And within our TEP, seven of the members of our technical expert panel had a smoking focus and among those seven five did

agree, either strongly agree or agree that this would be a good measure of quality.

Additionally, as you may have also read in our empirical validity testing, we found that smoking prevalence decreased by an estimated two percentage points for every ten percentage point increase in screening and intervention.

And so essentially when one does screen appropriately in one year we can see in the subsequent year that smoking prevalence does in fact decline, so there was an association there.

And the providers who performed well on the process whatever those processes are, in this case screening and intervention, they also performed well in the intermediate outcomes smoking prevalence, our measure.

One other point of this as an intermediate outcome measure we know that the measure is not required to be risk adjusted. That said, members of our TEP and our team firmly believe an unadjusted version of this measure is

the right path forward at this time.

First, we do not want to mask the underlying disparities that many of you have already articulated today that may exist within a provider's practice in terms of smoking behavior and among specific subgroups within their population.

Second, in support of broader population health improvement efforts we really do see this measure as an internal quality check essentially to say, assuming that you're utilizing the full scope or menu of options available to you whatever they might be, we can then determine and quantify the reach of those interventions by actually seeing whether or not your smoking prevalence has decreased as a result of employing those different options.

So we are still exploring the possibilities for the use of this proposed measure in our quality programs, but we are confident that it does offer critical linkage between discrete processes, so clinical

interventions, care coordination strategies and the like, and the downstream health impacts of smoking.

And we do also believe that it offers an opportunity, a very critical one, to potentiate the impacts of related process measures and to really provide additional flexibility to physicians as opposed to checking off just boxes around different processes to really be able to move us down to the outcome, or this intermediate outcome of interest and then eventually to our true outcome of interest which is reducing smoking related morbidity and mortality.

And so again just to echo what I said earlier, we believe that our measure's promise is really realized in first taking this initial critical step in focusing on population based outcome measures. We hope that it offers feedback to providers not only on the outcomes of the screening and intervention strategies, but also incentivizes them to leverage additional

approaches that are multimodal, that are evidence based, that are cross-sector, and most importantly that are patient centered.

I will stop there. I speak fast, but I wanted to make sure we had enough time for you all to weigh in, and we thank you again for this opportunity to have a discussion.

CO-CHAIR PINCUS: So initially are there questions about clarification of this measure that people had? So Jeff, Lisa.

MEMBER SUSMAN: I wonder what your response is to the fact that this potentially at least seems like it could be gamed. You're a smoker, I'm not going to bother screening you. You don't smoke, I'll make sure that you get in the sample and whether you have data.

I thought there was a fair amount of missing data, like a quarter or so in the population. I'm just trying to get a handle on that issue and I don't have any concerns about the validity or usability and so on.

DR. MCNAIR: I think it's a really

good question and I'll let Kate speak specifically to some of the data around the missing data. And what I will say is that's one of the reasons why we have talked about this being a really good opportunity to use in tandem. So assuming that with the other measure that someone has screened and is reporting on their screening practices, then this is a measure that could be used kind of harmoniously with that so that you could kind of get around to some degree.

It's a slight mitigation strategy around the potential gaming that could occur in that regard.

DR. SAPRA: Yes, thank you for that question. You're right about the missing data. And there's a couple of pieces that I wanted to highlight.

One is that when we looked at the full sample, so the testing was primarily done on those providers that had at least ten patients and screened at least 50 percent of all their patients, and then when we expanded it to all of

the eligible providers without those minimum criteria, the smoking prevalence was slightly higher.

So that does suggest that perhaps those people that are meeting the minimum reporting, do have minimum reporting for our testing did have somewhat lower prevalence. So that's something that we are open to talking about is whether or not there needs to be any kind of minimum reporting with this measure.

And then again as Dr. McNair said we would propose using this in harmony with the other process measures because if you don't screen someone you would not have them, because they are a smoker you would not have them in our measure and they wouldn't be in the denominator, but then for a process measure on screening they wouldn't be in the numerator.

So you would perform, let's say this is true of all your smokers you don't record anything on smokers but you record everything on the nonsmokers, you would perform really poorly

on a process measure around screening and you would perform great on our outcome measure. So it does provide a little bit of a check.

MEMBER SUSMAN: Yes, I'm not sure this is a feasible solution, but if you defaulted on the screened it means you're a smoker, there would be a high incentive in one measure to have everybody screened and to determine what their status was.

DR. MCNAIR: Precisely.

CO-CHAIR PINCUS: So Lisa and then Tami and then I have a question. Oh, and then Charles.

MEMBER JENSEN: I don't know if this comment fits in necessarily here and it's probably global for all of these tools that we're talking about tobacco, but I'm thinking about the legalization of marijuana in several states or a few states or however many states, where does that fit in, in our screening? Are we screening folks for use of marijuana?

DR. MCNAIR: I assume that, you know,

for most people that are taking adequate social history that they're including that. You're asking about tobacco, you're asking about illicits and you're asking about alcohol use.

But we don't have that directly included within our measure as is specified, but yes --

CO-CHAIR PINCUS: Sounds like that's another point too.

DR. MCNAIR: I think that's an important, I think that's a very important point to make, but I would hope that that's kind of how people are actually asking these questions.

MEMBER JENSEN: Well, it's not illicit in some states. So, you know, are you smoking tobacco? Yeah, no. But I smoke marijuana. I mean are they going to volunteer that? So just a thought.

CO-CHAIR PINCUS: So does tobacco. It's not a middle ground. Raquel.

MEMBER MAZON JEFFERS: Yes, I just had a question also about the care setting and it

relates to this issue about the preponderance of the smoking population being kind of in the behavioral health population.

So from what I can tell it doesn't look like the measure -- I'm with you. It doesn't look like the measure can be implemented in a behavioral health setting, or you don't specify that as a care setting for its use.

DR. MCNAIR: Right. Thank you for that question. And we do agree with all the points that were made and at the prior discussion about the importance of having this be a measure where we're able to focus on those populations that are most vulnerable, in particular those that are suffering from mental illness.

We actually don't explicitly say that but we do hope that it can be utilized in a behavioral health setting, outpatient as well as inpatient.

MEMBER MAZON JEFFERS: So just to be clear, it's people with mental illness and people with substance use disorders. And the amazing

thing about treating people for their tobacco addiction in a substance use disorder setting is that it's an addiction, so the people that are treating the other substance use disorders are very well equipped to be able to treat the tobacco use disorder as well.

CO-CHAIR PINCUS: So just to clarify, this measure would apply to all providers?

DR. MCNAIR: Are you talking specifically in terms of operationalization in the future?

CO-CHAIR PINCUS: Yes.

DR. MCNAIR: We're still determining that, but yes, I would say yes, our interest right now is for it to apply to all providers and I will tell you that we are thinking --

DR. BURSTIN: I'm sorry, I don't think Hal is asking about how you'd actually implement it but is the measure specified to be for all providers.

CO-CHAIR PINCUS: Yes, the measure is specified. That's my point. Yes, it's not how

you're going to ultimately use it, not how CMS is using, CMS is proposing this measure for endorsement, which means it's open for anybody to use for anybody.

So the specifications, you know, do not limit providers in any way, so potentially dentists, others.

MS. MARINELARENA: Hi, this is Melissa. Sorry, this is Melissa Marinelarena and I'm senior director here at NQF and I did help write the preliminary analysis and worked with CMMI.

There is in the specifications it does list the type of providers in there, so you have to look at the measure as the way it's presented to you right now. I believe there was, and Tracy, if you can help me. I think it was a psychiatrist, ophthalmologist, but it does, it's within the Excel spreadsheet.

But so right now it's only been, what's presented to you is the type of providers that are presented to you in the documentation.

CO-CHAIR PINCUS: So that's my question is what are those providers.

MS. MARINELARENA: And if you give me a second I'll, while you keep the conversation going I will pull it up for you.

CO-CHAIR PINCUS: Okay. I just want to clarify Raquel's question.

DR. MCNAIR: We were specifically saying that it is specified for behavioral health settings, but if you could pull up that list that would be helpful to us as well. Thank you.

CO-CHAIR PINCUS: Okay, so in the meantime, Tami?

MEMBER MARK: I thought I read that there's an existing measure parallel to this that applies at the state level, but you did think about applying this to the health plan level or why or why not?

DR. MCNAIR: NQF, yes, 2020 is at both the state and national level based on the CDC's behavioral risk factor surveillance system. And yes, this is something that could certainly be

applied at the health plan setting, at the health plan level, excuse me.

CO-CHAIR PINCUS: Okay, Charles.

MEMBER GROSS: A couple of questions to clarify. Medicare data only was used in the analysis?

DR. SAPRA: So I'll go ahead and take that one. It does use PQRS, but the EHR system which is actually all-payer, so for physicians who choose to report in PQRS EHR they have to report for their patients regardless of payer.

MEMBER GROSS: Thank you. And a second follow-up question, is the data captured here in some ways a subset of 3225 or should we hold, and if it is maybe we'll hold that for tomorrow?

DR. SAPRA: So our measure is very well harmonized with 3225 aka 0028, in fact that's where we took our measure from so it aligns very well. The big difference as Dr. McNair said at the beginning was that we are really concerned with the intermediate outcome of

smoking behavior.

So whereas our colleagues move on to, say, if you're a smoker then you need to go on and get either counseling or some other type of cessation intervention, we say we really just care whether you smoke or not and that allows providers, as Dr. McNair said in her introduction, to really use a whole suite of tools that are available to them that might not fit discretely within that particular process measure, or I believe we have another process measure coming up after ours.

So this one really works in tandem with all process measures, but it is extremely well harmonized and so hopefully this will come up when we talk about feasibility, usability, I can't remember which one. But it's coming right from 0028. Sorry, 3225.

MEMBER GROSS: Thank you.

CO-CHAIR PINCUS: Mike.

MEMBER TRANGLE: The focus of my question really is quite similar to yours in that

whether it's a subset or not it seems like this, if you're looking at the focus of the evidence, which I think is the section we're talking about here, it's not quite a standalone measure. You know, it's like if you only had this alone and it wasn't side by side with something else it wouldn't quite be meaningful and the focus would sort of be too narrow to be totally useful.

And to some extent I think I'm sharing your bias that is there a way of connecting or somehow mooshing together two measures into making one simplifying.

I know we like harmony, but even simplifying and combining as sort of a general principle for us to think about that and one that could really make it easier for the users and maybe less expensive. This is a prime measure to think about it.

CO-CHAIR PINCUS: Yeah, let me, also my question. Because when I first looked at this, I was one of the reviewers initially but I sort of looked through all of them. And I

initially thought it was something and the more I read it I realized it wasn't what I thought it was.

And so the relationship between a qualifying encounter and the relationship to this screening assessment, are they both one and the same thing in terms, you know, in terms of the time relationship between those two things?

So my initial thinking in terms of how I read it in a superficial way was that there was the creation of a denominator, meaning a denominator that was in that practice meaning somebody that you would be seeing, you know, on some potentially, you know, has had some contact with, and then at some later point to look at whether that person had remained, you know, had stopped smoking or remained a smoker.

But it looks like now it's simply, essentially a period prevalence of smoking, which means that if I, you know, if I'm a psychiatrist I'm going to have a much higher, you know, period prevalence than, you know, somebody who's an

ophthalmologist.

DR. MCNAIR: That's correct. I'll take this and then I'll see if Dr. Sapra has anything additional to add. You're correct that it is a kind of point in time estimate.

And just to go back to the definitions that essentially one would have to have had the qualifying encounter within the measurement period which is one year, but the ascertainment of this smoking status did not have to occur within that measurement year. Does that make sense?

So you would have had to have seen the patient within the measurement year, but the smoking status may have been obtained a year before that.

CO-CHAIR PINCUS: So I guess my point is that makes an attribution to me, so you're creating the attribution through the qualifying encounter.

DR. MCNAIR: Right.

CO-CHAIR PINCUS: But it does nothing

to say that, you know, that I did anything to improve things.

DR. MCNAIR: Right, I think that one of the, and you're right. One of the challenges is that this is a measure that will be very helpful over time, right, to essentially be able to trend over time and we recognize that. And we certainly thought about other opportunities to even refine the measure beyond what it is currently as a point estimate.

But that's why as we're proposing it today, we do think that it doesn't stand alone as well as when it is in tandem with another measure to be able to determine whether or not the things that you're doing at a practice level with your patients in terms of the different types of evidence based interventions and tools, whether or not they are having an impact.

CO-CHAIR PINCUS: So I guess I'm confused as to why you didn't propose this as a tandem measure, you know, because if it's not, don't all the endorsed measures have to be

standalone measures?

Well, Jeff, and well, actually, David was next.

MEMBER PATING: Well, my question's really along the same line because it all comes down to I'd rather do one measure than, why do two measures when one will work?

And I would be interested more that you would, I think that I'm not understanding the gap or the evidence for this particular measure which would make me want to stop you early, have you go work with PCPI and see whether you should be a component measure or a submeasure of their measure.

And I think with all the same kind of recommendations that Dr. Briss gave earlier, there's no doubt that looking at a panel based level or a population it may add something, but it doesn't seem like we should have a whole definition set to do that when we might already have data and can sub it out in various ways.

DR. MCNAIR: The one point that I'll

just make is that I think what becomes challenging about the proposition that you've made is that as what we're trying to achieve with our measure is that the breadth out of this process measure, I mean the menu, the suite what's available that can constantly evolve, right?

So what we don't want to propose is that you should always screen or intervene in this way. What we're saying ultimately is what we care about is the outcome or the intermediate outcome which is whether or not smoking prevalence has been reduced.

We're relatively agnostic as to how you get there, but I think the best way to get there is to utilize evidence based approaches.

MEMBER PATING: There's by taking smokers out of my practice, and I think we were also asking PCPI to evolve their measure as well. And so maybe these could be coevolved at the same time.

So rather than doing tandem to a

measure with a PCPI measure, which may need to be retired because it's not broad enough or not inclusive enough, maybe the two of you could work together to satisfy both these needs and evolve together.

CO-CHAIR PINCUS: So anyway let's continue. We have Mady, Racquel, and Tami, are you still, have your thing up? Okay, so Mady, Racquel, Tami, Jeff and Lisa.

MEMBER CHALK: So my comment is that this for me is a contextual measure. I don't know that we've ever talked in this group about contextual measures, but I know other groups have where you want to know what the prevalence is of something and then you set the other measures in that prevalence.

I don't know if this group has ever considered those as measures. Do you understand what I'm saying?

CO-CHAIR PINCUS: Yes, I understand. The question is whether that meets the sort of NQF expectations around --

MEMBER CHALK: That's my question.

CO-CHAIR PINCUS: -- measures that can be used for accountability purposes.

MEMBER CHALK: That's my question.

DR. BURSTIN: I mean in many ways this is essentially an outcome of smoking prevalence. So we do have, I know somebody's here from HealthPartners.

We do have an endorsed optimal cardiovascular risk measure for example that has four components, one of which is the percentage of nonsmokers in your practice. So I think in and of itself this is an outcome.

I'm not sure it's a component measure. It could be used as a component measure, but if you are going to look annually or compare practices to practices to see the prevalence of smoking.

Now that being said it certainly raises the risk adjustment concerns that many of you have brought up already.

CO-CHAIR PINCUS: Okay, Racquel.

MEMBER MAZON JEFFERS: So I really like the idea that you are moving towards an outcome based population measure and that, you know, essentially you're saying I don't really care how you get there, I just want to see your rate of population go down. I think there's real value in that and I support that you're moving in that way.

I think some of the concerns around the potential for gaming which sort of led into the discussion about trying to combine this with a process measure are important, right. So you said that practices were excluded from this study if they only, were only included if they screened 50 percent of their patient population.

Well, is it possible to really only include this measure for practices that are practicing universal screening in their practice so that then you have a different group of practices that you're looking at and you're avoiding some of the gaming issues that I think people are trying to, I think that's the

motivation behind trying to combine this with a process measure is trying to avoid this gaming issue.

But there might be a way within the measure itself to change the way you define the measure so that you can avoid the gaming within the measure so it stands more on its own two feet.

DR. MCNAIR: Thank you. That's an excellent point and echoes the sentiment that Dr. Sapra made earlier around trying to have a minimum kind of reporting threshold or burden that we would have to determine moving forward.

CO-CHAIR PINCUS: Tami, and then Lisa, Jeff.

MEMBER MARK: Hopefully this falls into the category of expanding and not beating a dead horse. So I think what maybe Harold is saying is it would be more helpful if you had a change measure.

So rather than just saying, because if you look at the numbers the range is, the 30th

percentile has six percent smoking and the 90th percentile has, let's see, 28 percent smoking, so if I'm in the middle why should I bother? I'm already doing fine.

But if the measure is based on change, then you're going to encourage everybody to do something not just people, not detail or, so I don't --

CO-CHAIR PINCUS: Lisa.

MEMBER MARK: I guess, yeah, like you had a response to --

DR. MCNAIR: Is it okay for me to respond?

CO-CHAIR PINCUS: Sure.

DR. MCNAIR: First of all, excellent point and I do think that a delta measure or something along those lines is certainly something to consider moving forward.

At this time we propose this, one, because we think that it again irrespective it does allow for this, again, quality check as we talked about as we think about it being used in

tandem with process measures.

But I do agree with you that a delta measure is also a direction that we could head in the future.

MEMBER SHEA: I just had a question and it might be in all of this about the attribution, because as an inpatient psychiatrist I might be the person who saw them in the hospital and they haven't seen anyone else and then they'd be attributed to me.

But I don't, other than the time I have them I don't have a chance to make any other impact. So I was just wondering, you know, we come into this too in terms of, you know, when you attribute who's the primary provider or how that works. So I was just wondering how you make that attribution.

DR. SAPRA: Thanks for the question. It's something that we have been talking about a lot with our partners across CMS, because I think actually this issue of attribution also gets a little bit into the operationalization piece of

it because this is specified as the patient panel prevalence measure.

So it could be used for an inpatient setting, it could be used in outpatient settings right now, and I know that Melissa and others were checking on this.

I believe that we have specified this as really for outpatient settings rather than an inpatient setting, but that's something that we're considering as we think about what is the appropriate venue to deploy the measure.

So that's, you know, we're trying not to bleed too much between the NQF endorsement and, you know, where this could potentially be used in CMS programs, but we are thinking a lot about these issues around attribution.

CO-CHAIR PINCUS: Jeff.

DR. LUSTIG: And just saying I, so under the care setting it says outpatient clinician office/clinic and then other, and then under other it says PQRS providers may include additional settings such as speech and hearing

evaluation, occupational therapy evaluation and ophthalmological visits.

MEMBER MAZON JEFFERS: And then also if you actually click on the setting, because I read that first and thought it excluded behavioral health. But then I found in the other document further, way, way, further down it actually specifies behavioral health settings.

MEMBER SUSMAN: So I think this is a great measure if, if we deal with the gaming. And we've already gone through two or three options to do that.

I think it's sort of like doing a clinical trial, a drug study where you have intent to treat and we measure actually what happens. And some people don't take the medicine, drop out, whatever, and we assign an outcome to them. We could do the same, you're a smoker until proven otherwise.

That's one option, whether that would be very acceptable, probably not. But the others that you proposed are also, but right now I'm not

going to pass this on validity because of that particular issue.

So I think as I'm thinking through this I see absolutely why you're doing this and where you're going. And ultimately the end game will be, we'll be responsible for populations, we'll be responsible for assuming that we are doing what we're supposed to be doing and held accountable for that. Great.

CO-CHAIR PINCUS: Mady, did you have yours up? Okay, Tami, yours up? Mike?

MEMBER TRANGLE: Just a quick comment or concern. And I think the concern, I mean this is like you said, sort of a point prevalence about how much people are smoking, but it's a point prevalence that has a two-year delay, window or delay potentially involved in that. And I don't know if your methodology is such it's the last one taken if you've have three or whatever it is, and I can kind of get a two-year delay if you're talking about screening, but if I think you're talking about prevalence it strikes me as

a little bit more concerning.

DR. SAPRA: So it is the most recent encounter. And the two-year is really to harmonize with our friends and really to reduce, I mean because we're very sensitive to burden and we did not want to be increasing provider burden with this measure. That's why we wanted to be fully harmonious with 3225 so that's the reason for the 24 months.

Now the last encounter that you have that's your smoking status. So if you were a smoker 20 months ago and then I saw you last month and you were a nonsmoker, that's wonderful for all of us and you would be counted as a nonsmoker in our measure.

CO-CHAIR PINCUS: So Rhonda and then Brooke, are you putting yours up?

MEMBER PARISH: Yes.

CO-CHAIR PINCUS: Rhonda then Brooke, and then let's stop and vote on evidence.

MEMBER ROBINSON BEALE: So I just needed some clarification with the 24 months

being there from the first encounter to the last encounter within that 24 months. However, are you requiring continuous enrollment or attribution to that particular provider for the 24 months?

DR. MCNAIR: Thank you. Are we requiring, so it's at the provider level but not from patient to patient. I'm not sure if I'm answering your question. Can you articulate it for me again?

MEMBER ROBINSON BEALE: Sure.

DR. MCNAIR: Thank you.

MEMBER ROBINSON BEALE: So if you're measuring me on a provider level, my population, and let's say, I don't know how you're going to do this with 24 months, but let's just say you have a measurement at the end of 2015, okay, and I have a 24 percent rate of smoking.

Then the next year I have a 55 percent smoking rate but I may have new patients that are attributed into my practice at this time in the next year, and I'm trying to understand how

that works in your measurement. And how do you account for that?

DR. MCNAIR: How do we account for like the turn --

MEMBER ROBINSON BEALE: Yes.

DR. MCNAIR: -- is that what you're asking? It's a very good question. Again because this is at the provider level, the idea is that you're going to have turn.

And so ultimately it's not that we're looking from patient to patient, we're looking at the level of the provider and of their panel. So that's, yes, that's my answer to your question.

MEMBER ROBINSON BEALE: But their panel can change.

DR. MCNAIR: Yes.

MEMBER ROBINSON BEALE: Unless you have continuous enrollment during that measurement period.

DR. MCNAIR: That's correct.

MEMBER ROBINSON BEALE: Okay, all right. You understand what I --

CO-CHAIR PINCUS: Okay, so Brooke, and then let's vote. Brooke, put on your microphone.

MEMBER PARISH: I just had a really quick comment that I noticed that this was smokers and nonsmokers, but a lot of the evidence was actually tobacco cessation which is also something different, and of course now we have nicotine cessation. Are we just looking at smoking and are we defining, is there any definition of smoking? Because I know that there are people who will say they're not a smoker and they have a cigar monthly.

DR. SAPRA: Yes, there we have specified what constitutes a smoker from the SNOMED CT codes and it is smoked tobacco, so it would include your monthly cigar. Or not you, maybe a partner or friend.

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

MEMBER GROSS: Hal, could you clarify, because we've been talking about a lot of different, what are we voting on exactly when it

comes to evidence?

CO-CHAIR PINCUS: So my understanding, and correct me if I'm wrong, whether the concept as presented is likely to be, is dealing with an important issue and is likely to have some impact on outcomes.

DR. LUSTIG: Right. Does the evidence base show that the measure is linked to outcomes.

DR. BURSTIN: This is an outcome so that's not quite right. So all we require for an outcome measure is that the developer has presented a rationale for the outcome and offers some information on potential evidence based processes or structures that could be utilized to try to impact that outcome.

So again different bar --

CO-CHAIR PINCUS: Now my interpretation is this is not an outcome measure, this is a prevalence measure. Outcome implies that there's a longitudinal context, to me.

DR. MCNAIR: It is an intermediate outcome measure. That's how it was submitted and

how it's specified.

MEMBER SUSMAN: A plan that does five percent versus 25 percent assuming we get the data issues would be better. Wouldn't you say that the smoking prevalence of you versus I if --

CO-CHAIR PINCUS: If it's a managed behavioral health plan as compared to an MCO I would think it was, you know.

MEMBER SUSMAN: That's what I'm saying, the level, but --

CO-CHAIR PINCUS: Yes, where at provider level you're comparing a community mental health clinic to a primary care clinic because it's --

MEMBER SUSMAN: And we always have those sorts of differences in these measures. There's always going to be different organizational structures.

CO-CHAIR PINCUS: It's not organizational, it's patient population.

DR. BURSTIN: Well, this is why I think this came back to them. We haven't gotten

to it, so again I want to try to separate out evidence from what will follow in terms of talking about validity and I think that's where issues like this of different patient populations and risks may come forward.

But I think for evidence again, if you think about this as an intermediate outcome measure the expectation is that there's a rationale for how potential processes for structures could impact this intermediate outcome or outcome.

MEMBER MARK: So what's the difference between evidence and validity? I'm getting a little confused.

DR. BURSTIN: Go ahead.

MS. JOHNSON: They are related, which is good, right? They should be related. When you think about evidence we usually ask you to think about the clinical side of things.

So for this particular one, do you feel like that there is evidence to show that if you basically stop smoking does that improve

patient outcomes? I think that's probably what we're looking at for this one.

You will talk about or could talk about evidence again later on when you talk about validity and that's where you think about how the specifications align with the evidence, if there's anything particular in terms of how the measure is constructed.

So I'm not sure if that answered your question but I can try again if that didn't do it.

CO-CHAIR PINCUS: So is everybody clear?

DR. BURSTIN: We're voting on whether stopping smoking is a good thing or not.

CO-CHAIR PINCUS: So it's not related to the measure, per se.

DR. BURSTIN: It's related to the measure focus. That's what evidence is. It's the evidence and the measure focus.

CO-CHAIR PINCUS: It's related to the intent of the measure.

DR. BURSTIN: Right, right.

MS. QUINNONEZ: Voting is now open for Measure 3229, patient panel adult smoking prevalence. Option, we are voting on the evidence. Option is number 1 is high. Option number 2 is moderate. Option number 3 is low. And option number 4 insufficient.

Voting for the evidence of Measure 3229. Option 1 high, option 2 moderate, option 3 low, and option 4 insufficient.

Thank you, all votes are in. Voting is now closed. For the evidence of Measure 3229, 39 percent voted for high, nine individual votes, 43 percent voted for moderate, ten individual votes, 13 percent voted for low, three individual votes, and four percent voted for insufficient, one individual vote. This passes the criteria for evidence.

CO-CHAIR PINCUS: Okay, so now let's talk about gaps. Any comments from the primary reviewers with regard to gaps? Jeff.

MEMBER SUSMAN: There is a nice table

demonstrating that there's a wide distribution of prevalences and that seems to me would, I'd dearly like to have zero percent and there are people in the 20s and so I think there's a gap.

CO-CHAIR PINCUS: Any other comments?

Okay, I think we're ready to vote on gaps.

MS. QUINNONEZ: We're now voting for performance gap of Measure 3229. Option 1 high, option 2 moderate, option 3 low, and option 4 insufficient.

Thank you, all votes are in and voting is now closed. For the performance gap of Measure 3229, 57 percent voted high. That's 13 individual votes, 39 percent voted moderate, nine individual votes, zero percent voted for low, and four percent voted insufficient. So for performance gap of Measure 3229 this meets the criteria.

CO-CHAIR PINCUS: Okay, now let's discuss reliability. Comments from the primary reviewers with regard to reliability?

MEMBER KNUDSEN: The liability was reported was pretty high, 0.89, I believe.

CO-CHAIR PINCUS: Other comments with regard to reliability, Jeff.

MEMBER SUSMAN: It's hard to separate out reliability from validity if the validity of the measure itself might be impugned, so that honestly you don't know what the prevalence might be if there's 25 percent missing data, which means that your measurement and my measurement might have a lot of unreliability in it or differences that are important to account for. So I think it sort of goes into both categories.

CO-CHAIR PINCUS: Okay, Shane, then Mike, then Racquel. Oh, Andrew, Shane. Shane, then Andrew, then Mike.

MEMBER SPERLING: All right, so just very briefly because I may have missed it earlier. This is self-attestation, right, are you smoking or not? Because I know the Safeway employee wellness program, they got a lot of attention in recent years.

Actually in order to qualify for that discount on your premium, you know, for your employer, your health insurance, they actually swab the inside of your cheek because there's some sort of test that actually tells you whether or not you've been ingesting tobacco or smoking tobacco.

DR. MCNAIR: Right, this would be upon screening and as reported by the patient.

MEMBER SPERLING: So self report by the patient.

DR. MCNAIR: That's correct.

CO-CHAIR PINCUS: Well, it's reported by the patient as reported by the provider.

DR. MCNAIR: Right, and there may be providers that are practicing other methods in which case they would also endorse that the person is a smoker.

CO-CHAIR PINCUS: Okay, Shane.

MEMBER COLEMAN: This might be related kind of to the gaming, I think it fits in this area. But and I'm curious if it's the necessary

sample size of the patients or the at least 50 percent screening I think was trying to get.

Are you guys just trying to get to make sure it's randomization? Because that's going to be one of the key things, right, to either validity or reliability is making sure that the screening that's occurring that you're grabbing are kind of a true random sampling of the population.

DR. SAPRA: Yes, these are, you guys are really hitting on really important aspects of the measure. And I do want to highlight that the reliability testing was done among those who had at least 50 percent of the patients screened and we don't know if it was a simple random sample of their patient population or not. We just simply don't have that information.

So I think this is one of the things that we've been talking about in terms of, you know, should this measure only be done for providers who are screening at least 50, or, you know, that's the way it was tested, at least 50

percent of their patients.

And when we apply that cutoff in a minimum of ten patients per provider we do get very high reliability. And so the reliability portion is really, you know, can we discriminate poor performers from high performers, and then validity is a little bit different, you know, are we measuring what we think we're measuring.

Are we measuring good and poor quality of health care that's delivered when we look at a very important predictor of morbidity and mortality which is the smoking behavior of the patient.

CO-CHAIR PINCUS: Okay, Mike and then Racquel.

MEMBER LARDIERI: Oh, yes. And I just have the same comment as with the other smoking measure and the reliability specifications to look at other settings, so I'll look at dental and include dental in there.

CO-CHAIR PINCUS: Racquel.

MEMBER MAZON JEFFERS: So to a point

that David raised earlier about the relapsing nature of the tobacco as an addiction is that, you know, if it takes on average seven times to quit and there's a provider who's working diligently with a patient who might be relapsing, it really does penalize that provider if that point in time happens to catch the patient when they have relapsed even if they've made several concerted attempts to quit.

DR. MCNAIR: I think that is a great point and I know that there was a comment from one of the reviewers around potentially adjusting for kind of the stage of change that providers -- oh, excellent -- that your patients are in.

I think again just going back to our earlier comment around the importance of just really unveiling kind of what you're working with, I mean that's really the spirit of this measure and we don't want to make assumptions and mask what we think is going on that may not be going on.

So that's really where we're coming

from with this measure.

CO-CHAIR PINCUS: I had a question about reliability also in terms of looking over a 24-month period. And you're making an attribution across that full 24-month period, would the sort of periodicity of different providers sort of create unreliability?

So if, you know, if some providers had sort of, you know, were more likely to do their screening at certain points in time and there was a sort of variation could that introduce reliability problems?

DR. MCNAIR: I'll turn to Dr. Sapra. We didn't find that in our testing but --

CO-CHAIR PINCUS: Because you're trying, I guess when you're trying to characterize the practice as a whole, but it's not really a point prevalence. It's a point prevalence at different periods of time over two years.

DR. SAPRA: It's the prevalence at the most recent encounter, so whatever the smoking

status is at that most recent encounter.

Now if you're someone who likes to screen your patients every two years because that's what the process measure, let's say you've been reporting on 0028 all along and so you screen every two years. That's the frequency with, and as long as you've seen a patient during our one year measurement period that's your periodicity.

Whereas, someone else down the road, maybe they're screening at every visit, and as long as they've also seen someone during the measurement year, both of you at the most recent encounter, wherever that status, whenever you ascertain that as long as it's within 24 months that is the smoking prevalence.

I don't know that there's a way for us to get around that periodicity because that's simply the way that it's specified and I'm not actually sure that it would affect the reliability.

I actually can't, maybe other people

can think through it with me, but from a measurement standpoint I'm not sure that it actually would affect the reliability of this measure.

CO-CHAIR PINCUS: Well, it might if you made changes to how you practice at different points in times it might, so anyway just something to think about. Yes, within that 24 months because you're not comparing people contemporaneously.

Okay, any other comments on reliability? Mike.

MEMBER TRANGLE: Just a quick question because I thought I heard this two different ways. When I look at the committee pre-evaluation comments I saw one of the comments was it does not include patients in a substance use disorder setting. But I thought I heard that it did.

So just, I'm a little confused about does it or not, and you might ask the same question if somebody's on a psych unit or in a

mental health setting too.

DR. SAPRA: Yes, it does include outpatient behavioral health.

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

MS. QUINNONEZ: Voting is now open on reliability of Measure 3229. Option number 1 high, option number 2 moderate, option number 3 low, and option 4 insufficient.

Looking for one more vote. Thank you. All votes are in and voting is now closed.

For the reliability of Measure 3229, 13 percent voted high, three individual votes, 43 percent voted for moderate, ten individual votes, 43 percent voted for low, ten individual votes, and zero percent voted for insufficient.

DR. LUSTIG: This ends up being consensus not reached for Measure 3229 for reliability, so we'll still continue.

CO-CHAIR PINCUS: So let's move to validity, 60 percent, so let's move to validity.

Jeff.

MEMBER SUSMAN: Yes, I think this is where this measure really falls down. We've talked about all the different concerns and I just don't think it's a valid measure as currently specified with all the issues that we've noted which I won't elaborate on.

CO-CHAIR PINCUS: Other comments of a different sort? Any comments supporting validity? Okay, I guess we're ready to vote.

MS. QUINNONEZ: Voting is now open for the validity of Measure 3229. Option 1 high, option 2 moderate, option 3 low, and option 4 insufficient.

Thank you. All votes are in and voting is now closed. For the validity of Measure 3229, four percent voted high, one individual vote, nine percent voted for moderate, two individual votes, 78 percent voted for low, 18 individual votes, and nine percent voted for insufficient which is two individual votes.

This does not pass the criteria for validity.

DR. LUSTIG: Now this is a must-pass criterion so our discussion on this measure stops.

DR. SAPRA: Can we ask for, I know we have discussed, but I just thought the precedent was that if don't pass that we can have some feedback specifically. Thank you.

CO-CHAIR PINCUS: People are, there are the particulars or recommendations in terms of feedback that people would like to make?

DR. MCNAIR: And we definitely welcome to the problems we've identified any potential solutions.

CO-CHAIR PINCUS: So I think, okay, I think two comments have been made so far that I think are probably the most significant and there may be others as well. One was to think about how this can be converted to a change measure.

Number two is to think about how this can be added to one of the previous measures that we've discussed that, you know, leaves it sort of over time to both screening intervention and

smoking outcome.

So that if it was sort of a combined nested measure it would, you know, be a more effective and potentially valid approach. And I don't know if everybody agrees with that or if there's other comments that people wanted to make.

So Racquel, Rhonda, other people?

MEMBER ROBINSON BEALE: Well, I would just say that there were two big concerns from my perspective. One was this gaming issue, so if there's a way to avoid, you know, providers performing favorably even if they're not screening their patients so, or because they're kicking all of their smokers out of their practice. So you really want to try to figure out how to control for that.

And then the other would be kind of the sequential nature of it. I think what people were trying to get out with the conversation around the window of time is that you want to see that the provider saw the patient, had an

opportunity to have an intervention with the patient before they're getting measured as to whether the patient is smoking or not.

So if there's a way to structure the measure so that it's more sequential in nature, which might get solved if you're looking at the change measure that Tami had suggested as well, the change in the performance, but so that the provider definitely had contact with the patient prior to being measured for whether the patient has changed their behavior or not.

CO-CHAIR PINCUS: Rhonda.

MEMBER ROBINSON BEALE: Yes, I just want to add to that. I think it's real important to look at your attribution methodology because it's real important. And I think you also have to have continuous and how do you say that the members continuously attributed to that provider during that time period. That is real important.

Otherwise, like if the provider gets a whole new set of patients in it's going to invalidate your measure over time or to show

improvement. So I think those are the pieces added to all the other comments would help to clean that one up.

CO-CHAIR PINCUS: Okay, anybody else? Okay, well, thank you.

DR. MCNAIR: Just want to say thank you all. We really appreciate this as we try to kind of push forward in this direction of outcomes. It's very helpful to have your expertise and to just hear your candid feedback. So we appreciate that. Thank you.

CO-CHAIR PINCUS: And so let's take a break for ten minutes. Promptly at 4 o'clock get back.

(Whereupon, the above-entitled matter went off the record at 3:51 p.m. and resumed at 4:02 p.m.)

MS. WILLIAMS-BADER: Hi, my name is Jenna Williams-Bader. I am director of Performance Measurement at NCQA. And I am joined by Mary Barton, who is vice-president of Performance Measurement at NCQA.

Today we are presenting the Medical Assistance for Smoking and Tobacco Use Cessation Measure. This is a long-standing health plan level measure that uses patient reported data from the CAHPS survey to assess if patients have received assistance to quit smoking and tobacco use from a doctor or other health provider.

It aligns with the Grade A recommendation from the United States Preventive Services Task Force, which I think the PCPI has done a good job. Are you describing?

The performance rates continue to demonstrate room for improvement, and we also have evidence that there are meaningful differences in these performance rates.

The measure has demonstrated validity. The questions undergo cognitive testing, and we also have case validity and construct validity. The measure also demonstrates reliability. And we've used the data binomial to analyze reliability of the measure.

And then the measure is used in

several programs, including the quality rating system, health plan ratings and accreditation, and the Medicaid adult core sets.

CO-CHAIR BRISS: That was mercifully brief. Thank you for that. So I think I'd like to suggest, we've talked a lot about a boatload of tobacco measures already. And we've generally decided, I think, that tobacco is bad. But there are things that can be done about it and that there's a remaining gap, right?

And so I think, unless somebody is dying to add stuff on the tobacco is still bad and things can still be done, that we could skip that criteria. And is there anybody around the table that wants to object to that?

And then can we -- and I think we have to -- we still have vote the remaining gap. But does anybody want to say anything further on the gap before we -- on minding the gap before we just move sort to a vote?

MEMBER SPERLING: And I promise -- I'm sorry, I promise to be very brief. I just want

to note for the record that, at least on the medication assisted on tobacco cessation, just in the last few months we've had some advancement. Because there was black box warning on one of the major products for people with psychiatric disorders.

An FDA advisory panel and the FDA itself voted to lift that black box warning, so a major, you know, evidence based and FDA approved intervention for smoking cessation is now going to be available to people with psychiatric disorders.

And I think it's something that would be important for us to try and get our hands around in terms of measuring how people with mental illness were able to access that FDA approved therapy.

CO-CHAIR BRISS: Congratulations. Anybody else have comments on performance gap before we vote?

(Off microphone comments.)

CO-CHAIR BRISS: Yes. We're done with

it. I skipped us right over the first hurdle. I've cleverly skipped that one and moved us straight to the performance gap.

DR. LUSTIG: But as a reminder, it's a maintenance measure, so since the evidence is in the same direction, we don't need to vote.

CO-CHAIR BRISS: And if you want to object on the evidence --

DR. LUSTIG: No, no, no.

MEMBER PINDOLIA: I just had a question. If you don't mind, could you just reword the CAHPS survey question for me. Like, word for word, I just can't remember it specifically. It's because of what's on the performance data. So that's why.

(Off-microphone comments.)

MS. WILLIAMS-BADER: Oh, I'm sorry. I didn't turn mine -- oh, okay. So you can hear me now. All right, so this is based, actually, on several questions. And I'm just pulling it up here.

Yes, if you actually look at, starting

on Page 101, it shows the questions that are used to determine the enumerator. And Page 102 shows the questions that determine the denominator. So for the numerator, as I said, there are three different rates.

So it's based on three different questions. "In the last 12 months, how often were you advised to quit smoking or using tobacco by a doctor or other health provider in your plan? In the last 12 months, how often was medication recommended or discussed by a doctor or health provider to assist you with quitting smoking or using tobacco? Examples of medication are nicotine gum, patch, nasal spray, inhaler, or prescription medications."

And the third question is, "In the last 12 months, how often did your doctor or health provider discuss or provide methods and strategies other than medication to assist you with quitting smoking or using tobacco? Examples of methods and strategies are a telephone help line, individual or group counseling, or a

cessation program."

CO-CHAIR BRISS: And the staff's getting real irate at both me and Harold for mixing up the criteria, right. And so I would say that how the measure is --

PARTICIPANT: What was that?

(Laughter)

CO-CHAIR BRISS: I'm sorry, sir. You're recused and don't get the rebuttal. We'll be here for the next three measure discussions. So I think the details of how it's specified is really a reliability thing. So first we should discuss, if needed, and then vote on whether there's a performance gap.

I would call specifications details a reliability thing. So anybody else want to comment? If not, let's vote.

MS. QUINNONEZ: Voting is now open for Measure 0027, Medical Assistance With Smoking and Tobacco Use Cessation. We are voting on performance gap, Action Number 1, high, Action Number 2, moderate, Action Number 3, low, and

Action Number 4, insufficient.

A performance gap of Measure 0027, looking for one more vote. Okay, voting is now closed. For performance gap of Measure 0027, 50 percent voted high, 50 percent voted moderate, zero percent voted low, and zero percent voted for insufficient. This passes the criteria for performance gap.

CO-CHAIR BRISS: So now we can move to reliability. So we've teed up how the measure was specified earlier. Would one of the lead discussants like to comment on reliability?

MEMBER MAZON JEFFERS: I just had a question, because the denominator is defined slightly differently in two different places in the PA. So at one point it says it's patients 18 years and older who answered the CAHPS survey in Medicaid and Medicare. And then another reliability specification, the denominator details are for Medicaid, Medicare, and commercial plans. So can you just clarify what the denominator is?

MS. WILLIAMS-BADER: Sure. You might have -- I don't know exactly what you've been looking at. But one of the things to note is that the advising smokers, smokers and tobacco users, the quit rate is only reported for -- it's only rate-reported for Medicare. Medicare does not include the discussing cessation medications and discussing cessation strategies.

The advising smokers and tobacco users who quit is reported for all three kinds. And then Medicaid and commercial also report the two discussing rates.

CO-CHAIR BRISS: Yes, ma'am?

MEMBER PINDOLIA: So my question is more to do with the wording of the questions themselves. And I know this is part of CAHPS.

Being one of the people that have to look at how we're improving our CAHPS scores and HOS surveys, one of the areas that I keep coming and running into, and then looking at the gap scores that I saw from 2014 to 2016, they're just pretty much always the same.

The concern I have is is the population really understanding what that means in the differentiation between all three of those questions? Because when we're doing our random samples, and then the feedback we're getting, I don't know if they really do.

So, like, when we have strategy or, like, could we make the language sixth grand level? I mean, could we do something so that, to make sure that it's not because of not understanding the question and that's why the results are coming this way?

Are we repeating it in so many different ways that they're kind of, like, I already answered this. Something to take back of understanding why is our measure continuously, year after year, not changing.

MS. WILLIAMS-BADER: Yes. So we do, all the questions do undergo commuted testing. So to your point, as much as we can, we use that commuted testing to help us determine whether or not individuals really are interpreting the

questions the way we intend them to.

I can't speak for certain as to why the rates haven't changed, but I will say I did go back and try to find if there's any other indications or other data that might align or not align with ours to show whether or not there are -- whether or not it's changing using other sources.

And there's data coming from the National Health Interview Survey, just related to the advice, that demonstrates that there has not been much change since 2005. So the rear deck where it actually got lower in 2010, but the 2015 rate for that is very similar to what it was in 2005. So we're demonstrating that there might actually not be much movement on this at the moment.

And that might be because patients are quitting, and so patients who are quitting are not in the measure. So we do know that individuals are -- fewer individuals are smoking now or using tobacco than they used to. So that

might be part of the issue and why the rates are staying stable.

CO-CHAIR BRISS: The other thing --

MEMBER PINDOLIA: That was my point, is that in the previous measures we're seeing some demographic data that shows, especially for Medicare, there's a continuous decline, especially as they get older. It's, like 1.8 percent, right?

So in our health plan, we just continue to keep getting -- keeping the same people. And they're just aging with us, right. So is our score -- because unfortunately, there's a lot of dollars tied to this, right, for 5-Stars and things like that.

So it does become, you know -- so then there's all this effort, like, why aren't we improving? And it's, like, well, I think we are. And so I just want to make sure we're not trying to say we don't have a higher quality when really we're not measuring the right people or the right way. And I understand it's a random sample, and

I get that, but something to really think about of what that means downstream effect.

MS. WILLIAMS-BADER: Sure. It think that's a good point and something for us to think about.

CO-CHAIR BRISS: The other thing that I would say on that point is that they say later that, when they do construct validity testing, that the different pieces of the three sort of things correlate with other.

And you would expect that, if some systems do better than other systems, which I find to be plausible, that people that reported more screening are also reporting more intervention. And that's actually what they show in their data. So that makes me feel a little bit better, that they're not just confusing people with the questions.

Comments on non-reliability? If not, why don't we vote?

MS. QUINNONEZ: The voting is now open for reliability of Measure 0027. Option 1, high,

Option 2, moderate, Option 3, low, and Option 4, insufficient.

Looking for one more vote. Thank you. All votes are in, and voting is now closed. On the reliability of Measure 0027, 23 percent voted high, five individual votes, 68 percent voted moderate, 15 individual votes, nine percent voted for low, which is two individual votes, and zero percent voted for insufficient. For the reliability of Measure 0027, this passes this criteria.

CO-CHAIR BRISS: Anybody like to volunteer to tee up validity, please?

(No audible response)

CO-CHAIR BRISS: I can do validity. They've done face validity testing with ten out of 12 supporting. They've done cognitive testing, as you've heard, and they've done construct validity testing, which surely you've also heard. So anybody want to really comment more before we vote on validity?

(No audible response)

CO-CHAIR BRISS: Let's vote.

MS. QUINNONEZ: Voting is now open for validity of Measure 0027. Option 1, high, Option 2, moderate, Option 3, low, and Option 4, insufficient.

Thank you, all votes are in. Voting is now closed. On the validity of Measure 0027, 41 percent voted high, nine individual votes, 55 percent voted moderate, 12 individual votes, five percent voted low, one individual vote, and zero percent voted for insufficient. For the validity of Measure 0027, this passes this criteria.

CO-CHAIR BRISS: Is there any further comment on validity before we move to feasibility?

(No audible response)

CO-CHAIR BRISS: Hearing none, somebody what to tee up feasibility rather?

MEMBER MAZON JEFFERS: The data is collected in a CAHPS survey which is feasible.

CO-CHAIR BRISS: Seems feasible on its face, doesn't it? Anybody want to add to that?

MEMBER SHEA: Well, just as a reviewer and NCQ report for the membership about confusion about it either.

CO-CHAIR BRISS: Okay. Any further comments before we vote? All right, let's vote.

MS. QUINNONEZ: Voting is now open for feasibility of Measure 0027. Option 1 is high, Option 2, moderate, Option 3, low, and Option 4, insufficient.

CO-CHAIR BRISS: And usability and use?

MS. QUINNONEZ: Looking for one more vote, actually. Here we go. Thank you. Voting is now closed for feasibility of Measure 0027. Fifty-nine percent voted high, 13 individual votes, 41 percent voted moderately, that is nine individual votes, zero percent voted for low, and zero percent voted for insufficient. So for feasibility of Measure 0027, this passes this criteria.

CO-CHAIR BRISS: All right, usability and use, anybody want to tee this up? Yes,

ma'am?

MEMBER MAZON JEFFERS: So I'm going for second place for the broken record club. So it's not clear whether this measure can be used in a behavioral health setting. And as we said before, tobacco use and the substance use in mental illness population is very high. So if you could just comment as to whether it can be used in a behavioral health setting.

MS. WILLIAMS-BADER: So again, it's not necessarily used in a setting as the other measures would be at the provider level. Because it is a health plan level measure.

The way that the question is phrased and the way the health provider is defined in the CAHPS survey is it can be a general doctor, a specialist doctor, a nurse practitioner, a physician assistant, nurse, or anyone else you would see for healthcare. So I think that depends on how the person taking the survey interprets that particular definition.

MEMBER MAZON JEFFERS: So I think it's

about 50 percent of states that still have a carve-out for their behavioral health program. It's not included in their managed care plans. And I believe, therefore, they're not required to complete the CAHPS survey.

I could be wrong. But I think it's a, you know, you're going to leave out -- you're potentially leaving out behavioral health providers.

DR. BARTON: So you're suggesting that this measure should be used in more places. And we could not agree more.

(Laughter.)

DR. BARTON: Unfortunately, NCQA does not have any pathway into measuring non-plan populations yet. But, you know, I think it would be -- a lot of this has to do with the sampling of per caps.

And of course that -- so that if there was a state that saw that its responsibility was to take care of these vulnerable patients whom they had not moved into managed care, they could

absolutely institute a survey of that population. I don't know of any states that have done that, nor does NCQA really have a lever to make states do that.

CO-CHAIR BRISS: Yes?

MEMBER SPERLING: While states may not have a direct lever to do that, they can. MBHOs are, in many cases, are compliant with NCQA standards, right? But while states -- while you don't have the leverage to do it, state can insist, in a Medicaid managed care contract, that you're caught up and will comply with all relevant NCQA standards, right?

CO-CHAIR BRISS: Yes. And Michael?

MEMBER TRANGLE: One, I agree with him. That's one of my two comments. But I won't repeat it. And the other is sort of the issue that in my efforts to keep first place, a program record amongst the population that is in your denominator, to be able to stratify it by behavioral health patients would be very, very important and helpful to do. So we could look

for improvement in that sub-population.

MS. WILLIAMS-BADER: So I would just say that would require, I think -- Mary and I were talking about this earlier -- that would require that they sample in a particular population. Because right now, the data that's captured through CAHPS would not allow us to stratify.

But actually, we see that the future of this measure is not a survey-based measure necessarily. But it might be a measure that uses electronic clinical data systems. So in the future, that might be something that we're able to consider. It's just not feasible with what we have from the CAHPS data right now.

CO-CHAIR BRISS: And Dr. Zun?

MEMBER ZUN: So one thing that I'm a little confused about is this is all contingent on the patient actually understanding that this is what they're getting, right, that they're getting this counseling or some intervention. Have we tested, in fact, that they know when

they're getting that intervention?

MS. WILLIAMS-BADER: So I think two things are not -- I mean, I can jump into it. One is, like I said, the measure, the questions you undergo on cognitive testing. But the second is that one reason to collect this from the patient is that the patient should understand that that's what happens.

And if they actually are saying no, this didn't happen, but the physician thinks that it did, that might actually be a major problem. So I think in this case, if the patient isn't understanding that's what's happened, then that might actually indicate a problem.

MEMBER PINDOLIA: I'll just add a comment of clarification on that. Les, I'm not sure if you're aware, but the survey is done, like, April through June, like, in 2017, for them to remember what happened in 2016, right.

And so I don't know if you can really say that for many of our Medicare patients. They're 80 years old. They're on many different

medications. And sometimes they don't even remember what happened a month ago, much less a year ago.

MEMBER ZUN: Well, that was one of my concerns, will they know about it.

MEMBER SHEA: So meanwhile, I mean, and I don't know whether you could do this, is if you could bring claims on your own for the people that indicated that they had or didn't have smoking cessation with medicine, or if you bring a sample to see if they filled it or didn't fill it.

DR. BARTON: We have not done that in the past. A survey, you know, one of the downsides of doing a survey measure is you don't know who the people are. CAHPS closely guards the anonymity of their sampling.

But as we think about a future measure that Jenna referred to, I think we would absolutely be interested in looking at something that triangulates data from prescriptions, from claims for counseling, or claims in places where

this is done, you know, things like quit lines, in order to assess, really, what's been provided to patients who still smoke.

CO-CHAIR BRISS: And I wanted to make two quick comments. So one on that point, neither of your options for this are perfect, so there is -- you can either -- your realistic choices are a provider checkbox measure that isn't accurate all the time, or the patient's report of whether I actually got the service.

And the truth is if I had to choose between these, I would pick the patient report which I think is closer to outcomes, right. And so -- and then the other thing that I wanted to say that nobody has said yet is that these measures are, in terms of usability and use, there isn't a whole lot of programs. Anybody else? Mike, are you still --

MEMBER LARDIERI: Well, yes. I just want to back up a little bit. But there are 41 states that recognize NCQA health plan accreditation. So it's only nine states where

you wouldn't have that.

PARTICIPANT: D.C. is one of them.

(Laughter.)

MEMBER LARDIERI: D.C. is one of them. Sorry, DC. But it covers most of them.

CO-CHAIR BRISS: So can I move us to a vote?

MS. QUINNONEZ: Voting is now open for usability and use of Measure 0027. Option Number 1 is high, Option Number 2 is moderate, Option Number 3, low, and Option Number 4, insufficient information.

Looking for one more vote. Thank you. All votes are in. The voting is now closed. For usability and use of Measure 0027, 45 percent voted high, that's ten individual votes, 50 percent voted moderate, that's 11 individual votes, five percent voted low, that's one individual vote, and zero percent voted for insufficient information. So this actually -- usability and use criteria is passed for Measure 0027.

CO-CHAIR BRISS: We'll talk about the whole portfolio of tobacco measures tomorrow, so we don't have to do this today which, I think, gets us to the overall suitability vote. And anybody have final comments before we vote?

(No audible response)

CO-CHAIR BRISS: Hearing none --

MS. QUINNONEZ: Voting is now open for overall suitability for endorsement of Measure 0027. Option Number 1 is yes, Option Number 2 is no.

Looking for one more vote. Thank you. All votes are in. Voting is now closed. For the overall suitability for endorsement of Measure 0027, 100 percent voted yes.

CO-CHAIR BRISS: All right. So as I discussed, we're going to move the discussion of the portfolio of tobacco measures to tomorrow which makes us back some time. And so two more measures for the afternoon that we're going to try to get through. So without further ado --

(Off-microphone comments)

CO-CHAIR BRISS: That's right, 108.

DR. LUSTIG: So we were conferring here before we move on to the next measure, NQF has a new designation for measures that exceed our expectations for the criteria. And so Measure 0027 appears to qualify for us to consider whether it should get what we call endorsement plus designation.

And this is the first time we're actually going through this process which is why we wanted to confer about this. And so you'll notice it does say, toward the end of our measure evaluation form, that this measure is a candidate for endorsement plus designation if the committee determines that it meets the evidence for measure focus without an exception, which this did, is reliable as is demonstrated by score level testing, is valid as demonstrated by score level testing, and has been vetted by those being measured or other users. And so our initial analysis was that the measure did meet all of these criteria.

So really, it's a discussion of the group whether we think that this should be conferred this endorsement plus designation.

MS. JOHNSON: And, Tracy, you may want to just remind everybody about the vetting and what we're looking for for vetting. And actually, Mary, you may want to talk a little bit about what data comes in to you guys. Because the first piece has to do with giving data back to plans. And you guys might not give it back, they might give it to you. So you might just need to describe that for us.

CO-CHAIR BRISS: And can I ask a question of the staff too? So what are the implications of endorsement plus? So if I'm a user, and I'm trying to make some sense out of the population of 600 or so NQF fit endorsed measures, how am I supposed to use this information?

DR. BURSTIN: So some of it was the idea that these would be measures that exceeded our criteria. And in particular, we added a new

criterion for this particular designation which is that you seek input from those who are using your measure to actually improve upon your measure.

And so some of this was trying to get more of that usability into the development process. And so wanted to reward those measures and developers that were actually, in fact, doing that. And it would give an indication to others out there that this is potentially a measure that, you know, achieves the highest rating from NQF.

MEMBER TRANGLE: I have a question, and I hope it doesn't sound too metaphysical. But it really gets to if we're a behavioral health measurement standing committee, what does that mean?

And I think it kind of gets to are we looking at the -- is our domain sort of behavioral health patients wherever, whenever they come up? Is it broader, is it all humanity? And we just -- some of the measures may apply

more to us. You know what I mean.

Because if we take it that we're really sort of stewards of how well quality of care, and satisfaction, and other things are going on with behavioral health patients, it might play out differently where we can't really subdivide them here. And do we want to give a special endorsement versus our domain is larger than that? So I'd just like us to at least think about that?

CO-CHAIR BRISS: Raquel?

MEMBER MAZON JEFFERS: Yes. So it's hard for me to endorse plus something when I'm still struggling with this harmonization issue. Because this measure feels -- I know you said that it's different from the measure, whatever it was, 3225. Because it's not a provider checkbox, it's a -- I get the difference.

But they're really measuring the same thing. Did patients receive tobacco screening and cessation advice? And so it's hard for me to vote to endorse plus something in a -- maybe I'm

not thinking about the process correctly. But it's hard for me to think about endorsing plus something without taking into account the other similar measures.

CO-CHAIR BRISS: So, Helen, is there an answer to why this is really different from the 3225? It sounds like the answer might be that NCQA has gone through additional effort to get input from users. Is that the short answer?

DR. BURSTIN: That would be my sense of it. But again, I will also admit to Rachel's point that we didn't specifically think about whether the absence of a competing measure would preclude you from this plus designation either. So you've raised a really good point. Again, these are different data sources. But it's still a fair point.

CO-CHAIR BRISS: Mady?

MEMBER CHALK: Yes. So, Raquel, so to me the difference between the two measures is one is collected by providers or plans and the other is a patient survey. To me, one is more patient

centered, and the other is not.

I'd love to see somebody, some researcher, get those into a relational database and be able to take a look at how responses differ and how that matters. But that's a separate issue. That's a separate issue from the measure itself. And the fact that you're getting feedback and asking for it is very important to me, from patients.

CO-CHAIR BRISS: Helen, has NQF endorsed -- done endorsement plus before this committee, or are we guinea pigs again?

DR. BURSTIN: No, actually, I wasn't aware we were actually rolling it out yet. It was mainly rolling out the new criterion that says did you get feedback, are you iteratively improving the measure, which was really our action for now.

But, you know, I don't know that we have to spend a lot of time on this. And we can return to the plus designation after we've had more time to deliberate. But nonetheless, just

so pleased to hear that, on the surface, you guys read that. So that's great.

CO-CHAIR BRISS: I'd like to table this discussion, because I'm afraid it's going to hang us up. And the thing that I'd like to have you guys really work out is how to handle the plus designation in the context of a family of measures down the road. I'm a little worried about unintended consequences for some measures of the family and not others. So with that --

DR. LUSTIG: So I think we're moving onto Measure 0108. And we do have two committee members that are recused. Connie and Harold are both recused from discussing this measure.

DR. BARTON: And if it's okay, Dan Roman will introduce the measure.

MR. ROMAN: Hi, I'm Dan Roman. I'm a senior research associate at NCQA. I'm going to be discussing our follow-up care for children prescribed ADHD medication, so switching gears a little bit from smoking.

There's a long-standing plan level

process measure that uses medical and pharmacy claims data as its data source. Our intent here is to ensure appropriate follow-up care occurs once children are prescribed ADHD medications to allow for assessment of medication effectiveness and adherence, and to monitor for potential side effects.

The measure applies to children six to 12 years of age who are prescribed ADHD medications during a 12 month intake period. There are two rates. The first, the initiation fees, assesses follow-up within 30 days after the first dispensing event.

And then the second is our continuation maintenance phase which assesses, of those who are compliant for Rate 1, how many then have two additional follow-up visits in the 31 to 300 days after the first dispensing event.

Since the measure was first introduced in HEDIS set in 2006, on average, across commercial and Medicaid plans, performance rates have increased six to ten percent for the

initiation phase rate and ten to 16 percent for the continuation and maintenance phase.

We also see a wide variation in performance across the percentiles overall. So we believe there is still an opportunity for quality improvement. And as far as changes, we have not made any since the measure went through its last maintenance update in 2015. That's all.

MEMBER PINDOLIA: So I'm going to read my comment, because I put a lot of thought into this one on the details.

So regarding the maintenance measure and the new information, there's been a large shift in patient provider interaction to follow more technology-based methods, whether it's either EMRs, the chats that go along with that, Apps, direct emails to physicians, telemedicine.

This measure limits the initial follow-up visit as face-to-face within 30 days of starting the ADHD medication.

The second change in patient care that's really growing rapidly is the high amount

of high deductible plans that are rolling out. So this has a direct impact on the patient's out of pocket spend that has increased the physician's desire to find alternative methods to have patient communication, and especially for the young adults, which is the six to 12 year-old parents, where if their child is fine, and they did that first visit, they're really reluctant to come back within 30 days and pay another office visit, which is out of pocket for many of them with their high deductible plans.

So with that, I'm asking for evidence on the first initial phase of the 30 days. I understand it comes from the 2011 AAP guidelines. So I've actually contacted AAP, because there is no literature to support that statement. AAP has responded that they agree, there is no literature to support that statement, but the body of individuals in 2011 felt it was appropriate.

And I said in 2011 it probably was. But healthcare has evolved so much. And now, five, six years later, that is not the way we're

practicing. In fact, we're telling our providers you've got to find different ways to engage with your patients. This whole coming back in the clinic is not working.

So then, I also went and did some guidelines of lines of my own. And in January of 2017, there was actually a study published in American Journal of Managed Care. And it was actually assessing this actual measure. And it was done through the Medicaid population in Alabama in 1999 to 2012, 61,251 patients on average per year.

And their findings also really help share the concerns that I have that the low performance that we're seeing in that first 30 days, because there was a stark difference between the two follow-ups, the 30 percent goes to 60, 70 percent, their study showed that actually the ones who were the poor performers for the 30 days, within 30 days, had higher -- no, had lower ER use and hospitalization.

Because the patients who have those

biggest concerns, the parents are willing to bring them in. So the ones who did have their follow-up within 30 days, they actually had higher ER and hospitalization use. They had higher total cost of care compared to those who were not compliant.

When they went and expanded that day period, 30 to by just 20 days, you see a 20 percent increase, and improvement, and compliance to that measure. But, you know, so at this point, I don't know what that initial phase is really telling me, that that's poor quality within 30 days.

The heart rate and the blood pressure increase that, you know, people are worried about, I think, from talking to AAP, that's why they had this, the parents bring the children in after -- in the late evening. The peak effect of the drug has worn out. So to catch that, you're missing it.

So I've actually reached out to the Henry Ford Medical Group chair of Pediatric

Psychology, Psychiatry, and the chair of Pediatrics for their input on this measure. And they said exactly, that is their concern.

And this has become a checkbox for them and to just have them come in and have a nurse check it, meaning nothing. Their biggest concern is one to two months later, when they're down to one percent BMI, and that's like, their biggest issue they're having with the startup of the ADHD drugs.

And they are all being given titrations, right, for the first month anyway. So we're overburdening these parents with so many outreaches and then telling them to come back, but they're not doing so. So I have a lot of concerns of the evidence to support that first 30-day visit of being face-to-face and within 30 days.

MR. ROMAN: As far as it being face-to-face, you're correct. It is based off of the language in the 2011 guideline which suggested that there should be a face-to-face.

That said, NCQA is evaluating the use of telemedicine and telehealth in general across all of our measures. For this particular measure, we have out for public comment a recommendation to include, I believe it's video conferencing for this measure.

It already does include telehealth for the other two visits, so we are evaluating that and looking into it to try to make sure that we're counting, exactly, everything you just said.

It's kind of in process though, right now. So the change is not implemented in the measure under the version that you're seeing. But it is something that we have out for public comment right now.

MEMBER PINDOLIA: Yes, when would that change occur if -- because it has a great implication on the providers. Because it goes right into so many provider fee incentive payments.

MR. ROMAN: It would -- I believe it

would be implemented for HEDIS 2018. So that's published in July if it goes through public comment, and our CPM, our Clinical Performance Measurement, approves the change, and it goes through our Board, and all that.

But July is when we come out with the next version of the measure. So it would be another way for qualifying for the measure beyond the face-to-face.

CO-CHAIR BRISS: Presumably, that would be a substantive change that would require, for it to maintain endorsement, it would come back through us again, right? Is that right?

DR. BARTON: Of course we would update the specification. And that's actually a decision about whether ---

(Simultaneous speaking.)

DR. BARTON: We would share the updated specification with NQF. And I believe they would decide whether or not it merit a separate review.

MEMBER MARK: On the point of evidence, the summary document says that the

developer states numerous, "More than 100 studies related to the care of patients with ADHD have been published since the publication of this guideline. None of which contradict the need for appropriate follow-up once treatment with medication begins."

But it wasn't clear if you did any kind of systematic review, or how you reviewed the evidence, or what went into that. I mean, my little, you know, simple review came up with some interesting articles, you know.

And things continue to change, as was pointed out, in this field, like, if there was any study on how consumers felt about having to come back in, if they felt burdened. That might be helpful.

The rate of prescribing of ADHD drugs has increased greatly. So it'd be helpful to have more information about exactly what was done in terms of the updated literature review.

MR. ROMAN: Sure. I would say it would not rise to the level of a systematic

review. For a maintenance update like this, we do look at the studies available and do kind of a cursory review of is there anything that is saying this measure is outdated, or is no longer effective, or doing harm as follow-up.

And so as far as follow-up goes, we were not able to find anything that says that there's -- definitely you shouldn't be doing follow-up in an appropriate timeframe. I mean, that's really kind of the level of things we look at for a maintenance update.

If there's a guideline change, for example, we start there. In this case, there aren't any guidelines. Even the AAP, though they said that they, you know, that they don't really think that face-to-face visit's necessary, there's no new guideline out.

So we start at the guidelines, and then we kind of work our way down into the other literature. And there is plenty about care for ADHD on children. As far as anything contradicting the need or the appropriateness of

the measure, we were not able to find anything.

MEMBER EINZIG: So a few points. First point is I'm really impressed you actually got in touch with the American Academy of Pediatrics.

And just to emphasize, you know, guidelines are not evidence. And so if it was just a bunch of people around a table saying this is what we think is right, that's not evidence, not strong evidence. So I want to emphasize that.

Second thing, if you pull up the AACAP practice parameters, they don't give timeline. I don't know if you're able to put it up on the big screen. Can you put up the link for that AACAP practice parameter?

I'll just read it off here. "The frequency and duration of follow-up sessions should be individualized for each family and patient, depending on the severity of ADHD symptoms, the degree of comorbidity of other psychiatric illness, the response to treatment,

and the degree of impairment in home, school, work or peer-related activities.

"The clinician should establish an effective mechanism of receiving feedback from the family and other important informants in the patient's environment to be sure symptoms are well controlled." And it goes on.

They suggest two to four follow-up visits in a year. But nowhere in it does it say a one-month timeframe in the first month. So that's the second point.

The third point, just from a clinical perspective, from seeing a lot of patients, I can come up with a lot of examples when it is not in the patient's or family's best interest to come back in within a month.

The first example would be prescribe in July, school doesn't start late August or September. It might not be the most fruitful time to come back before school starts. You won't get optimal feedback.

Another example is you've got a kid

with autism who also has ADHD symptoms. And that can be a huge burden for an emotionally disregulated kid to have to come in for a follow-up within that timeframe. So I can go on, but I'll stop there.

CO-CHAIR BRISS: So it sounds to me like several people have expressed concerns about the level of evidence supporting the one-month timeframe. So -- and the truth is the last time the committee met we had some of these concerns even then.

And as has been said already, our tools for interacting with families have only gotten greatly better since the last time we revisited this measure.

So let's -- I'd like to suggest that we call that issue sufficiently aired, unless somebody has new points to make on that. And then we'll go the rest of the way around the table. Please, Mike, you're next.

MEMBER LARDIERI: My question is more about how you're going to implement the video

conferencing. And not all plans pay for it. If you're in a Medicaid environment, some states don't allow you to do that. So you're sort of stuck, for awhile anyway.

MR. ROMAN: I can say a little bit. And let me see if Mary or Junqing, who is also here, Liu from our NCQA, can add more. There is a coding system structure that's out there that is going to be implemented. And we put that into our specs.

As far as state by state, I'm not sure, you know, how much control that NCQA has over what is allowed. But there are codes available that can be used for video conferencing and, for Pela Health, and modifiers to existing codes that capture this sort of thing, the kind of the more electronic exchange.

MEMBER LARDIERI: Yes, I'm aware of the codes, but how do you get providers paid? I mean, it's great to have the code, but you're not going to get paid for it. So you're not going to -- I don't know a lot of them are going to do it.

MEMBER ZIMA: I just have two additional, unique points. One was, you know, this measure's been around for ten years. And I kind of disagree that there's been a lot of change. So, I mean, I think it really raises the question of really, you know, how valuable has this measure been stimulating change over the last ten years?

And I think it's particularly important, because in the CMM phase there's always that potential overestimation of adherence. Because there's a very high attrition rate, right. So what, 70, 80 percent of kids don't make it into the CMM phase. Is that right?

(Off microphone comments.)

MEMBER ZIMA: Sixty-five percent. And I think the other issue is I think we have to be mindful that these are Schedule 2 drugs required in triplicate, except for atomoxetine.

And, you know, both AAP and AACAP guidelines really emphasize that, in monitoring medication safety in these children, you have to

look at the child clinically. You have to look at the side effects, you have to look at the parent preference, you have to look at things like school time, and school schedule, and things like that.

And, you know, when you look at the fine print in that CMM, it appears that two of the follow-up visits, one is allowed to be just a telephone contact. And, you know, to me I don't think a telephone contact with a parent who wants a refill of a stimulant in a hyperactive kid, and I haven't been able to see that child, check the pulse, check the blood pressure, check the weight, I don't think that's safe practice. And so I think those are two of my main concerns.

And, you know, I think the other issue is it gets back to, you know, over 100 publications supporting continuity of care. Maybe it's continuity of care when you could set Rx persistence that should be measured as well during the CMM, not just the number of contacts.

MEMBER TRANGLE: I think these are

related, but somewhat unique points. When we talked about this last time, I think we also talked a fair amount about what is the literature showing about starting and somehow adjusting stimulants as to whether true outcomes, functional abilities to do well, and adapt in various social and academic settings, what's the evidence for that versus just that we got them relying on a pill, you know?

We said that was sort of -- we talked about it last time and said it wouldn't hurt to see that burnished and just make sure there's good correlation there.

And the other thing, my other comment is, at least in my system, we're doing a lot of communication in other vehicles that you didn't mention. And I'd hate for us to modify it to televideo only so that we can stay behind where things truly are at by the time it gets implemented.

So we're doing a lot of sort of just online, going through My Chart, you know, kind of

directly into the EMR kind of communication. We're doing mobile apps. We're doing other kinds of vehicles that you did not mention but are becoming more prominent, you know. And, you know, I just don't want to always stay behind the curve.

MEMBER PATING: Yes, my comments are actually to the same. And I want to nominate Vanita for next Senator or President.

(Laughter)

MEMBER PATING: Also for going the extra mile when I was trying to even just to get the first 100 feet.

You know, NCQA has done a really great job with moving the field in a lot of different ways with these initiation continuation formats. I think you've done it with this. You've done it -- I believe with depression. I know that you've done it with alcohol and drug screening.

But the problem is is that the field is evolving faster than you all. So the measures are supposed to drive performance. But I

actually feel in my system, which is a large networked system, NCQA has actually inhibited, through its auditing process, the ability to innovate.

I mean, one is televisits. There were video visits. Another one is apps, another one is even just various kinds of things, including education which is not considered billable appointments, because of this data structure that you've created.

And to have another two years of this, you are putting institutions out of product cycles. That is a whole product cycle in the technology world, in terms of being able to use IT hand-held technology to monitor people that can give you very good outcomes.

So organizations are not willing to take the risk, because NCQA auditors in the field will not, you know, modify specifications, you're telling us that it really starts from the top. And so I'm just really asking you to hear the comments around the table. And it's not that

you're not moving the field in a good way, you have moved the field. But the field needs to move you, and it needs to be reciprocal, right.

So you just need to hear that in a very active way. Because I think you've put my organization two or three years behind by not willing to innovate the measure. And it's shut down innovation in very important ways.

CO-CHAIR BRISS: Okay. So we've heard a lot of kind of similar comments around the table. I think we may be at a point where we can vote on evidence unless somebody else would like to have a last word. And the question on the table is essentially do you believe that the measure, as currently specified, has evidence to suggest that it importantly drives better outcomes.

MEMBER MARK: Well, I think -- that's the validity measure. This is the generic question of whether this is an important -- if there's evidence to support that, in general, right, following up after ADHD treatment is

important.

MEMBER PINDOLIA: Is it that, or is it the evidence of what we're trying to measure with this?

MEMBER MARK: I think that's the validity. Again, it gets complicated and confusing several times --

CO-CHAIR BRISS: Yes. We've been a little back and forth. And I'm not sure that there's an easy answer to that.

MEMBER MARK: Yes.

CO-CHAIR BRISS: I think this is the key question. I think you could make it an evidence question. And I think you could make it a validity question. I heard from you, it would have been an evidence question. But I'm okay if we limit our conversation about evidence to is ADHD important. And, you know, is continuity of care important?

MEMBER MARK: Yes. So the only other thing I would say is that the recent report from SAMHSA reports that the number ED visits

involving ADHD stimulant medications tripled from -- I'm trying to get things updated here -- tripled over the last ten years. So in 2011 it was 31,000 visits. So seems like there are a lot of increasing harms being done by ADHD medications which might argue that this is more evidence for this than there was before.

MEMBER PINDOLIA: If I could just add a caveat, I think you have to look at that per thousand. Because the number of children that are added on for ADHD over the last ten years has increased so much. So the increase in ER isn't just as proportional as the increase in the number of people using the drugs.

MEMBER MARK: No, you're right. That's my point. There're so many more people using the drugs. We're having a tripling of the adverse effects. So that would argue that we need to have post to start follow-up even more now than we did before, not post to start, you know, post-prescription follow-up.

CO-CHAIR BRISS: All right. So I

don't actually think it's worth voting on your ADHD. It is important in whether follow-up conceptually is also important. So I'd actually like to vote on the evidence question about do you think that this measure, as currently specified, is likely to result in meaningful improvements for affected kids and families.

MS. QUINNONEZ: Voting is now open for measure 0108, Follow-up Care for Children Prescribed ADHD Medication. We're voting on the evidence. Option Number 1 is high, Option Number 2 is moderate, Option Number 3 is low, and Option Number 4 is insufficient.

Thank you. All votes are in. Voting is now closed. For the evidence of Measure 0108, five percent voted high, so one individual, 40 percent voted moderate, eight individual votes, 35 percent voted low, seven individual votes, and 20 percent voted for insufficient, four individual votes.

DR. LUSTIG: So that's consensus not reached, and we'll continue to discuss the

measure.

CO-CHAIR BRISS: So we move to reliability. Any other reviewers like to comment on performance gap? Anybody like to comment on performance gap?

(No audible response)

MEMBER ZIMA: I think it's been said, little change.

CO-CHAIR BRISS: And then having said that, it is little change, and there are probably important differences, at least as the measure is currently specifying, between bottom performers and top performers.

MEMBER ZIMA: Yes. So the bottom 10th quartile is 29 percent, and the 90th top quartile is 50 percent for commercial insurance. And for Medicaid it's 29 percent and 56 percent. So there's a big gap in performance.

CO-CHAIR BRISS: So anybody want to talk about that further before we vote on that criteria?

(No audible response)

CO-CHAIR BRISS: All right, let's vote.

MS. QUINNONEZ: The voting is now open for performance gap of Measure 0108. Option 1 is high, Option 2, moderate, Option 3, low, and Option 4 insufficient.

CO-CHAIR BRISS: But if it's a hung jury, you keep going. We had a hung jury.

MS. QUINNONEZ: All votes are in. Voting is now closed. For performance gap of Measure 0108, 33 percent voted high, six individual votes, 50 percent voted moderate, nine individual votes, 11 percent voted low, two individual votes, and six percent voted insufficient, one individual vote. For performance gap of Measure 0108, this passes this criteria.

CO-CHAIR BRISS: So we're done. We move to reliability. Comments?

(No audible response)

CO-CHAIR BRISS: We've clearly worn out the panel. So they've done a signal to noise

analysis. It meets usual standards, it looks like. Anybody want to comment further?

(No audible response)

CO-CHAIR BRISS: Why don't we vote.

MS. QUINNONEZ: Voting is now open for reliability of Measure 0108. Option 1, high, Option 2, moderate, Option 3, low, and Option 4, insufficient.

Thank you. All votes are in. Voting is now closed. For the reliability of Measure 0108, 35 percent voted high, seven individual votes, 40 percent voted moderate, eight individual votes, 25 percent voted low, five individual votes, and zero percent voted for insufficient. For reliability of Measure 0108, this passes this criteria.

CO-CHAIR BRISS: And then we're to validity. And so it's both components of that. Does it measure what it purports to measure and, if you do the right things that people will be better off, right? So, Bonnie, you're up.

MEMBER ZIMA: I think, again, this

gets to David's point about the NCQA approach. I mean, this will be based on the face validity, you know. And you look at the technical expert panel, only three of the panel members were MDs or prescribers of medication. And of those three MDs, only one was a child psychiatrist. One was an internist, and one was a family medicine person specializing in geriatrics. There was no pediatrician on the initial TEP. So I think that should be called out.

And then the other thing, looking at construct validity, you pulled up another HEDIS measure related to any contact with a primary care provider. And I guess that makes sense, because you're really looking at them at a follow-up visit.

But again, I think the more important issue is the continuity of care for children on stimulant medication. And so I think you kind of missed the mark a little bit in thinking about construct validity there.

CO-CHAIR BRISS: Anybody else have

points they want to make that haven't already been made?

(No audible response)

CO-CHAIR BRISS: Let's try voting this criteria then.

MS. QUINNONEZ: Voting is now open for validity of Measure 0108. Option 1, high, Option 2, moderate, Option 3, low, and Option 4, insufficient.

Thank you. All votes are in. Voting is now closed. For the validity of Measure 0108, ten percent voted high, two individual votes, 25 percent voted moderate, five individual votes, 55 percent voted low, 11 individual votes, and ten percent voted insufficient with two individual votes. This is consensus not reached.

CO-CHAIR BRISS: No, it's --

MS. QUINNONEZ: Not passed?

CO-CHAIR BRISS: Not passed, right.

MS. QUINNONEZ: No pass. For the validity of Measure 0108, this is a no pass for validity.

CO-CHAIR BRISS: Thank you.

DR. LUSTIG: And so we do stop our discussion on this measure at this point. We do have a post-meeting comment period which, if you want to submit any other information, you can for that.

CO-CHAIR BRISS: I think we were pretty clear about what the issues were. We've given other developers a chance to ask any follow-up questions. Do you have left-over questions about what the issues were?

DR. LUSTIG: No.

CO-CHAIR BRISS: Thank you. So 0576, we are in the home stretch.

MEMBER TRANGLE: May I just ask you a question? I mean, whether it's now an effort, what happens now. I'm not familiar with what happens if you choose to deconstruct or not have a -- is there a lag of time period that it just continues to happen, do you see what I'm asking?

DR. BARTON: NCQA's not dependent on NQF endorsement. We use our measures. We're

delighted, overjoyed to have the input from the fields that this process gives us. We think there's a super-high value to the consensus development process.

Don't let me -- big mistake is not respecting the CDP process. It's wonderful, it's terrific. But just as CMS, sometimes uses measures that are not NQF endorsed, sometimes NCQA uses measures that are not NQF endorsed.

CO-CHAIR BRISS: But it does seem likely that the feedback I'll get, you'll get it incorporated into your ongoing discussions, right?

DR. BARTON: Yes, absolutely.

MEMBER COLEMAN: Can I just check in real quick? The summary of each of those actions doesn't come up for me. I'm just curious if any others have noticed that too and make sure there's not a technical challenge on my behalf that I need to fix or something.

(Off microphone comments)

MEMBER COLEMAN: Correct. Yes, I've

noticed they haven't been on other ones, but now that it's mine, I just want to make sure that I don't need to do something different.

CO-CHAIR BRISS: All right. So we are on 0576. Do you guys want to tee that up for us, please?

DR. LIU: Sure. This is Junqing Liu, research scientist at NCQA, delighted to be here to talk about the Follow-up After Hospitalization for Mental Illness Measure. This is a long-standing claim-based measure. The measure addresses follow-up care for a vulnerable population who is hospitalized for mental illness.

Evidence shows that follow-up care reduces suicide attempts, re-admissions, and improves functioning. Clinical practice guidelines also support follow-up after hospitalization.

The measure has demonstrated performance gap in a variation across health plans. About 30 to 40 percent people across private lines do not have follow-up care within 30

days after discharge from hospitals. Even evidence supports it is important to receive follow-up care.

The measure has high reliability as demonstrated by the beta-binomial reliability scores. The measure is used in seven national reporting programs such as Medicaid, a coreset, hospital compare data in 70 programs, in-patient psychiatric facility quality reporting programs. And the measure is also used in NCQA's accreditation of commercial, Medicaid, and Medicare plans. Thank you.

CO-CHAIR BRISS: Thank you. Anybody like to go first and talk about the evidence for this one?

MEMBER COLEMAN: I'm happy to start us out. Again, I only have my comments here, but I'm happy to share them and others can chime in with me.

I think the biggest question for me came up around the seven and 30 days. Again, that's kind of the time period. I wasn't sure how

much evidence there was specifically for those two time points. So I was curious why they were chosen.

Otherwise, I wanted to make sure it was psychiatric hospitalization. You mentioned that, hospitalization for mental illness. It actually, I thought I saw it in two different applications, or two different parts of the application it reads just a little bit differently. So I just wanted to clarify that. And I think that's mostly it.

CO-CHAIR BRISS: We'll go around this way. Raquel?

MEMBER MAZON JEFFERS: I just had a question that with our efforts to really try and develop more and more integrated models of care, even for people who might be slightly more moderate even to acutely mentally ill, it looks like the follow-up visit can only occur with a mental health specialist, mental health practitioner.

But more and more people are receiving their mental health services in primary care

settings as well. And I'm just wondering if you thought about that as an alternative option for people.

MEMBER MAZON JEFFERS: She's waiting for more questions?

CO-CHAIR BRISS: I think you can go ahead and answer that one.

DR. LIU: Thanks for the comments. So the seven-day and 30-day, those are consensus based. We developed this measure with our Behavioral Health Measurement Advisory Panel. The panel considered that if someone is sick enough to be hospitalized, it's important for them to receive timely follow-up care.

So we felt that it will be great if someone can receive follow-up care within seven days. If not, at least they should receive follow-up care within 30 days. So that's the rationale behind those timeframes.

And I think studies have started to emerge using this measure. It is the -- and started to show that follow-up within these

timeframes are contributing to reduced re-admissions.

In the psychiatric hospital measure you mentioned, actually this measure is about hospitalization in any hospitals. It could be psychiatric hospitals or other general hospitals.

And your comments about integration of primary healthcare and considering other types of providers, we thought about that. And actually we think the integration was not contraindicated with this measure. Because most of the integrated models, first you need mental providers either available remotely or being part of the team.

And we also, consulting some research conducted, such as by the World Health Organization research, is showing that still in primary care half of the time in the mental health program conditions are not recognized.

Even when they are recognized, half of the time they're not treated. For those who are treated, in the majority of time they are not the state of art first line treatment provided by

other type of providers.

So we discussed with our Measurement Advisory Panel in great detail. They felt that, again, someone who is sick enough to be hospitalized, they should see a mental health provider.

MEMBER MAZON JEFFERS: I don't disagree that there are many models of primary care behavioral health integration that utilizes part of their workforce mental health professionals to do a lot of the mental health interventions.

But there's a prescribing component. Would you -- I guess my question is do you consider a physician who might be a prescribing physician who's prescribing a psychopharmaceutical for someone with a mental health disorder, do you consider that a follow-up visit?

DR. LIU: Yes. If someone has a prescribing authority, yes.

MEMBER MAZON JEFFERS: And that's a mental health professional.

CO-CHAIR BRISS: So let's keep going

around. Tammy, I think you're next.

MEMBER MARK: Yes. Again, I think the -- well, if you can talk about how you updated the literature, so I'm thinking back to that last discussion. You know, it seemed like the way it went was, if there was nothing contradicting this, our prior evidence was still good. And mainly they relied on guidance, you know, looking at practice guidelines.

And if there were studies that they had cited, that showed that ADHD follow-up led to improved outcomes, that might have changed the conclusion of that last discussion. So I'm wondering if you found any, or you looked for any articles that had looked at the correlation between this HEDIS measure and some outcomes, like re-admissions. Because I think that would be helpful.

Because we do seem to be really thinking about the validity again. And so if there's newer evidence in support of the validity, it would be helpful to also hear about that.

DR. LIU: Yes. About the process of evidence review for this measure, we actually tried to be very comprehensive when we do those runs. And we also listened to internal work looking at this measure.

So we tried to be as comprehensive as we can. You can notice that we have several updates in the form about practice guidelines and evidence since last endorsement. We had a study cited after that last endorsement.

And we at NCQ are thinking hard about how to correlate our measures. I think that's a discussion about how we can correlate this measure with our Plan All-Cause Readmission measure.

At this time, that measure does not stratify the condition. But we can look into it. And as I mentioned, there are published studies conducted outside of NCQ, is using this measure demonstrating the measure can reduce re-admission.

CO-CHAIR BRISS: Another thing that might be at least subtly different from the last one, as I hear it, is for the last one there were

lots of other proposals that were made about other ways that could assure adequate follow-up that weren't actually captured by the measure, right.

And we did that in both directions There were ways that people proposed that weren't captured by the measure that might have been less face-to-face. And there were times when, later in the course of the telephone visit, might not be enough.

And so the other thing that, at least ever since the discussion at this point in this measure, is that we haven't had the same kind of discussion about there are better alternatives to assuring follow-up.

MEMBER MARK: Yes. I mean, I published a study showing that post-discharge follow-up reduces re-admissions. And it wasn't cited, so I know there's literature out there wasn't cited. So my point is that it could be helpful to, you know, bring that forward more clearly.

(Laughter)

MEMBER LARDIERI: Yes, thanks. I see

the, you know, the list of providers. But how are you picking that up in the claim on the primary care side, like, that it's a social worker, or that it's a masters or doctoral degree professional? How are you picking that up on the claim?

DR. LIU: We received questions from the field through HEDIS medical reporting. People ask us how they should identify the right type of providers. We heard that there's NPI and other providers, emergency providers indicated.

Our claims and our HEDIS management specification also provide definitions about mental health practitioners. So there's other ways to help plans' providers to identify the mental health practitioners.

MEMBER ZIMA: Yes. You know, I think I had a question and a comment. One was this measure goes down to age six. But why didn't you report any evidence supporting it for children and teens?

DR. LIU: So I think this measure we

were talking about in children and adults. We actually are looking into evidence, and there is evidence supporting the prevalence of this condition. I think this panel can all agree that mental health conditions is also prevalent in children, adolescents, and adults.

I think we can -- if that's, you know, something we need to add, we can definitely add more. Because we are looking into that. It is important for that population as well.

MEMBER ZIMA: Yes, I thought it was an important omission, because the evidence was basically from NICE and then APA by target disorder, schizophrenia, bipolar, and major depression. So that, I thought, was an important omission.

The other issue is really this conundrum that we talked about that, you know, we're taking measures where the unit of analysis is the health plan. But when it gets NQF endorsed, then we're putting hospitals in a position of being accountable for this measure.

And, you know, NCQA can argue, well, you know, when we had a health plan level measure, the health plan is accountable to find follow-up for their patients. But as I can imagine, using this measure for a hospital, they're going to face the challenge of coordinating follow-up mental health care in a system that's fragmented, or their contract to the commercial insurer that doesn't have the capacity to accept all of their referred patients. So I think that, in using it prime time, it kind of shifts who we find is accountable.

DR. BARTON: So I would absolutely take that up with someone who proposed this measure for the IPF population, for IPFs. And that's not what we're proposing here. We're proposing this as a measure for health plans.

MEMBER KELLEHER: And it has been used by at least the, we know from the carve-out managed behavioral health organizations, it's been sort of gold standard that they themselves do. They don't rely on the provider network, including

the facilities, to follow-up, the seven and 30 has been pretty standard stuff for a very long time.

CO-CHAIR BRISS: Absolutely.

MEMBER KELLEHER: And very useful for reducing re-admission.

CO-CHAIR BRISS: Just before I continue around, it seemed like people were coming out of the woodwork to want to comment on this specific thing. So anybody have further comments on this specific issue? It's sort of about the level for which its specified, right? At least --

MEMBER SHEA: So I can tell you that what the health plans do is they use this as a quality metric for the hospitals and they take money away from us if we're not performing, if our discharges aren't performing. So that is definitely happening now.

CO-CHAIR BRISS: Okay, Rhonda.

MEMBER ROBINSON BEALE: I can testify to that.

(Laughter)

MEMBER ROBINSON BEALE: To the fact

that we use it for both hospitals and the outpatient facilities, generally the IOPs, because it does bring a joint accountability. So I think limiting this to the health plan is really not as effective as it is beginning to apply it to the delivery system itself.

But one of the things that's been inherently problematic with this particular measure, and behavioral health has adopted telemedicine quite extensively in some areas, and you're still not allowing that to be part of the follow-up which I think is still problematic at this point. A visual, audio visual contact with the patient can be very, very sufficient in behavioral health.

CO-CHAIR BRISS: So Lisa wanted to get back in that one.

MEMBER SHEA: Yes. I have another -- it's kind of like a flip, but sometimes plans want the hospital to, like, set up an appointment that very day. So they leave the room, and they walk across the hall. So they're outpatients now. So

I didn't know if you have data about same-day visits of discharge versus after that. Because that's a strategy that has been used.

CO-CHAIR BRISS: Hang on, hold on a second. You guys want to answer that question?

DR. LIU: Sure. So I think this kind of is very well acknowledged that this kind of measure is challenging, right. We always write accountable entity hospital outpatient health plans.

Our thinking is health plans are paying for the services, they are responsible for the network accuracy across inpatient and outpatient. So we see health plans that involve accountability entity.

But we know these measures are being used for hospitals. We think this is encouraging care coordination. Because all the entities are in the same loop. We need -- I think, I know places are working together to improve the performance.

But the Powerhouse issue, we are

evaluating Powerhouse for all HEDIS measures. For this measure, we are adding video conferencing for the follow-up visits. And if that is approved, we will have our annual updates for NQF process. We update our measures all the time, any changes, you know, specification, we want to make sure we are aligning the version that's available for use, that's endorsed by NQF.

About the same-day visit, we heard of that. And we're trying to understand the magnitude. We think that's a small portion. We actually had a discussion with our Behavioral Health Measurement Advisor Panel on that very issue.

And the recommendation came out of that panel is that that's great quality improvement effort, that places are trying to catch people while they are still in your care to, you know, make that appointment.

But that's not equating the clinical care of a follow-up visit. So for that we made a decision to balance the challenge of reporting and

the quality of true clinical interventions. We decided to remove the same-day visit. This is a, you know, another recent decision. We wanted to make sure next round of annual update of this measure, we want to make sure that's clear.

CO-CHAIR BRISS: Mike, are you still on this point?

MEMBER TRANGLE: You know, we have the same-day sort of gig going on by a number of health plans in our area. And when I talk to them about this, and sometimes it felt like there was confusion running rampant about is this a continuity of care measure where the ideal is you're getting to see your provider you're going to continue to see on a routine basis, and you've made a transition, versus I'm hooking you up with someone who you're going to see just as an interim kind of thing, tide you over while you're waiting a couple of months to get in to see somebody else.

And they both meet your measure. But I think the things you're really trying to incense are quite different, you know. And some of that

depends on how much access, in general, you have in your area.

CO-CHAIR BRISS: So generally speaking, panels wear out as the day goes on. You guys are getting more riled up as the day goes on. Shane, were you --

(Laughter)

CO-CHAIR BRISS: Shane, were you trying to get in?

MEMBER COLEMAN: No, I'm --

CO-CHAIR BRISS: Les, you're next.

MEMBER ZUN: I'm really concerned that we're giving the wrong message here. If 50 percent of psychiatric patients have a medical illness, and between 20 and 80 percent have a substance use disorder as well, and we're sending them back just to a mental health professional, I think that's the wrong message.

I really think we -- and your issue about integrative health is much more important. I don't know if I can say any philosophical or ethical kind of comments here today, but I think

this is -- you know, although I don't have a problem with the measure and what we're trying to do, but we're going in the wrong direction here. As a national organization that's trying to be on the forefront of quality, this is not where we need to be.

Okay, I'm off my soapbox. Thank you.

MEMBER ROBINSON BEALE: Okay, last statements, you need to include substance use disorders in this. I'm not sure why that's continued to be excluded. That's a huge part, and it's also an are that is exceptionally important that you have follow-up.

FEMALE PARTICIPANT: Besides, it's a mental illness.

MEMBER ROBINSON BEALE: And it's a mental illness, according to S05.

(Laughter)

MEMBER ROBINSON BEALE: Maybe. But the last point is this measure would be far more effective if you also had a composite measure that measured engagement post-discharge. Someone who's

been in the hospital is going to need to be seen more than once and probably three times in the next 30 days or so.

So a composite measure will help stop the gaming of the same appointment, the same-day appointment by the hospital. Because a person is going to have to see the next level of provider and demonstrate some engagement.

CO-CHAIR BRISS: So, Mike, I'm going to give you the last word before we vote this criteria.

MEMBER LARDIERI: Okay. I was going to echo what Les was saying, because there's a lot of that going on out there where the plans will pay just for someone to meet the patient in the community someplace for a visit to get their seven-day visit. And then they shuffle them off to who their real provider is. And they may not get to that provider for awhile. So they met that 30-day.

And then I have the same comment on the video conferencing. I'm all for video

conferencing. But if you're going to approve it, you need to get the plans to pay for it and work it out in the states so that in the Medicaid plans -- so that you can actually do it and get paid for it. Because that's not happening now across the states. It's not a code, the codes are there. It's state laws and some plans are not paying for it.

CO-CHAIR BRISS: So Bonnie wants to edit me and get the actual last word.

MEMBER ZIMA: Sorry, last minute. But, yes, I'm a little concerned because of the severity of the illness of the patients that are being hospitalized for mental illness and the breadth of the mental health providers spanning people that aren't able to prescribe.

I would anticipate that a substantial proportion of the patients being discharged are going to be on some type of psychotropic medication. And so somebody could pass and get a counselor, which is good, but a big reason for re-admission is that they're not compliant with their

medication.

DR. BURSTIN: This has been an incredibly interesting discussion. It's gone way beyond evidence. So I get it. And you'll have, you know, if you continue it, other opportunities to bring that in and other points along validity, or usefulness of the measure, or feasibility. But again, this is truly just about did they provide sufficient evidence for the measure focus. I just want to keep us back oriented to the task at hand.

CO-CHAIR BRISS: So with that, yes?

MEMBER ZIMA: Just for clarification, so our mental health participant is being defined as a psychiatrist, a psychologist, a social worker, not the primary care provider, if that's who's been the continuous provider for that patient?

DR. BURSTIN: It specifies mental health provider also include social worker, registered nurse, marriage and family counselor, professional counselor.

DR. LIU: Yes, that's the definition.

It does not include primary care physicians.

MEMBER MAZON JEFFERS: Like you said, you're going to be reconsidering video contact, and you're looking at some other updates to the measure. Is this a consideration for -- are you considering updating the measure with a new definition of provider, community provider?

DR. BARTON: I'll say the most immediate changes are to the codes for telehealth, and types of telehealth, and the same-day visit exclusion. The question of the provider, the way that we specify who the person should meet with, is a longer term question that we're continuing to engage with our panels.

I think, you know, Bonnie's raised the point that the very sick population who are discharged from mental health hospitalizations, and our advisory panels have continued to hue to the sense that it should be a mental health professional, maybe not a prescribing professional but at least a mental health professional who can understand the reason why they were hospitalized

the issues that are likely to come up as a result of their discharge.

And so that's, really, the advice that we've taken in terms of specifying this as being a visit with a mental health professional.

CO-CHAIR BRISS: So I'd like to move us to the vote. And is there -- so now we're talking about the evidence for the measure concept, as it's currently written.

MS. QUINNONEZ: We are now voting on Measure 0576, Follow-up After Hospitalization for Mental Illness. We are voting on the evidence. Option Number 1, moderate, Option Number 2, low, Option Number 3, insufficient.

Thank you. All votes are in. The voting is now closed. For the evidence of Measure 0576, 75 percent voted moderate, 15 individual votes, 20 percent voted low, four individual votes, and five percent voted insufficient, one individual vote. So for evidence of Measure 0576, this passes the first criteria.

CO-CHAIR BRISS: Anybody want to

comment on gap in care?

MEMBER COLEMAN: I'm happy to keep us moving. You know, it suggests basically that there is a gap in, you know, looking at coverage for Medicaid and Medicare to commercial coverage. The average rates basically, like, with this measure, where anywhere from 30 to 50ish percent, 90th quartile went from -- or looking at ten to 90th, you're looking at anywhere from, you know, 30s to 80s as far as one of the largest spans, it looks like.

And then there's, you know, there were remarked to be differences, you know, disparities among the usual things, such as age, SCS, I think we said, social functioning, minority status, et cetera.

CO-CHAIR BRISS: And for the record, there is not such a thing as -- there are, like, four quartiles. There's not such a thing a the 90th quartile. I think there is a 90th percentile.

All right. So anybody else want to comment on gap? Yes?

MEMBER MAZON JEFFERS: I just had one question for the gap. Do you do analysis to see if there's a difference between rural and suburban areas or inner city?

And it goes back to the conversation we had earlier identifying other individuals that could possibly help with the health plan being responsible. We really are trying to find other ways to help our providers, especially in the areas where they can't reach. And so it may be even telemedicine might not be an option. Because we have patients that just don't even have a computer, you know, in the home and things like that.

So could a health plan behavioral health person help satisfy, and connect to that, and do things? So just to think about it from a different perspective.

CO-CHAIR BRISS: So, Bonnie, were you trying to get back in on this point, or is your card -- okay. So it looks to me like we can vote on performance gap.

MS. QUINNONEZ: Voting is now open for performance gap of Measure 0576. Option 1, high, Option 2, moderate, Option 3, low, and Option 4, insufficient.

Thank you. All votes are in. For the performance gap of Measure 0576, 40 percent voted high, eight individual votes, 60 percent voted moderate, 12 individual votes, zero percent voted for low, and zero percent voted for insufficient. For performance gap of Measure 0576, this passes the performance gap criteria.

CO-CHAIR BRISS: So on the reliability front, I think we've talked about the details of the specifications in some detail already. So let's not re-litigate a lot of the specs discussion that we've already had. Would anybody like to tee up the reliability testing or new things about specs that we haven't talked about already?

(No audible response)

CO-CHAIR BRISS: Are we ready to vote?

MS. QUINNONEZ: Voting is now open for

the reliability of Measure 0576. Option 1, high, Option 2, moderate, Option 3, low, and Option 4, insufficient.

Thank you. Voting is now closed. For the reliability of Measure 0576 we have 30 percent voted high, six individual votes, 55 percent voted moderate, 11 individual votes, 15 percent voted low, three individual votes, and zero percent for insufficient. For reliability of Measure 0576, this passes this criteria.

CO-CHAIR BRISS: So that takes us to validity. Anybody have additional comments that haven't been made? Bonnie?

MEMBER ZIMA: Just to give -- face validity is based on the NCQA measure life cycle. And advisory panel's meaningful difference is determined by statistical differences between the 25th and 75th percentile ranking.

CO-CHAIR BRISS: David? David?

MEMBER EINZIG: So this is just going back to defining mental health providers. So places that have collaborative care models, the

first question is if the psychiatrist gives advice to the pediatrician, but that's not accounted for in the billing or captured in some way, that person will get dinged. So that's Question Number 1.

And then Question Number 2, for pediatricians or folks in primary care who get extra training in mental health, their pediatric portals -- I don't remember what it's called -- but there's extra training that people can get for a year to get more mental health expertise or ongoing six-month with follow-ups with a child psychiatrist or a mental health professional. Will those folks be counted, or do they get captured in some way?

DR. LIU: So the provider type, I understand we have a recovery care model is that the primary care and then the psychiatrist or the mental health providers, they can both bill the services. So therefore, that service will come through our measure if one of the mental health providers' service is captured.

The other part about the pediatricians who have extra training about mental health, I think that's moving in the right direction. Because there is a shortage of mental health providers, especially child psychiatrists.

So we -- our measure, we can see how much advantage, and research is demonstrating, and how the states' definitions of the mental health providers will evolve. We will definitely want to make sure our measures keep pace with those developments.

CO-CHAIR BRISS: So, Dave --

DR. BARTON: I would just add to that that the additional training for pediatricians is absolutely in the spirit of our list of providers that are accepted. And so that's, you know, I think -- and what Junqing said is absolutely right. So we need to work on figuring out how we can word things in order to make sure that we're sorting the right people in.

But to your first question, I think, really, the truth of the matter is that claims

data is certain on the one hand, because everybody wants to get paid for the thing that they do, but on the other hand, practice is moving very fast in ways that suggest that claims data is not going to be the best way to determine the truth.

And we're very actively working on measures, including a set that is not coming to this panel, about monitoring for depression, that use electronic clinical data that is completely unrelated to claims. And I think that that's among the ways in which the future is going to look.

CO-CHAIR BRISS: David, Michael, Shane.

MEMBER PATING: I would just like to, again, reiterate that the expansion of the type of providers that can be providing services with behavioral health homes and medical homes, I think the likelihood of primary care providers being the medicater or the point of contact is very significant, as well as in many states Medicaid only allows one visit type per day.

So you're either going to see your internist, or you're going to see your therapist.

And if I have a heart failure, I am going to see my internist. And it's not going to fall in. So it should really be any provider, any contact with anyone can improve the initiation and engagement, including unlicensed substance abuse counselors.

CO-CHAIR BRISS: Michael, Shane

MEMBER TRANGLE: My comments are along the same lines in that there's great variability in terms of whether PAs are covered by health plans. And there are collaborative care codes. I don't know when they take effect. But I don't think they specify which patients are talked about, you know, so that I could, whenever it takes effect, I could say I had a discussion with primary care, but you won't know what patient it's about, you know.

So I think ultimately you're looking at EMRs and trying to figure out -- it would be in the chart in the progress note that talked to Dr. Trangle about so-and-so. So I think that's a good way to go.

CO-CHAIR BRISS: So it's pretty clear

at this point, I think, that we've established that there are -- there's an ever-broadening range of providers that you guys might consider. So, Shane, I'll give you what I think will be the last word on this one.

MEMBER COLEMAN: Yes. So I guess I'll keep it short. The care codes, they came back in January, they tie in to the PCP visit. But you can find, like, a G-code that says the psychiatrist was consulted.

States haven't activated the Net, but they exist in seamless code. So definitely there's a -- and then otherwise, I'll just -- I want to give a little bit of what we call impose in my organization.

And I totally agree with the spirit of everything everybody's saying. But if we have any primary care docs here, I think we have, like, one, right, maybe? Yes. They freak out sometimes if you send schizophrenics to them out of the hospital unsupported.

We're in a really innovated system. So

I just want to say, yes, I actually agree. I totally agree with the spirit. And, I don't know, just be aware that, you know, you start to be a little bit -- when folks have severe and persistent mental illness, you absolutely freak primary care out if you start, you know, discharging them out of a psychiatric hospital straight to primary care without any sort of help.

CO-CHAIR BRISS: So it looks like all the cards are down. I'd like to try to vote validity please.

MS. QUINNONEZ: We are now voting on validity of Measure 0576. Option 1, moderate, Option 2, low, Option 3, insufficient.

All votes are in, voting is now closed. For the validity of Measure 0576, 63 percent voted moderate, 12 individual votes, 37 percent voted, low, seven individual votes, zero percent voted insufficient. So for Measure 0576, validity, this passes the validity criteria.

CO-CHAIR BRISS: So that takes us to feasibility. Anybody want to tee this up? This

--

(Laughter)

CO-CHAIR BRISS: Yes, this one's already in play, right. It seems feasible. Anybody want to comment on it further before we vote? Let's hear it please.

MEMBER KELLEHER: Oh, okay.

CO-CHAIR BRISS: Oh, sorry. Did I jump the gun?

MEMBER KELLEHER: I just wanted to say that, being from one of the nine states where the behavioral health system is entirely not integrated with the physical health system, it is still challenging in many places to have someone discharged from a hospital setting and caught in the community, and to have those data systems talk to each other sufficiently to gather this measure.

CO-CHAIR BRISS: Anybody else have comments?

(No audible response)

CO-CHAIR BRISS: And if not, let's vote feasibility please.

MS. QUINNONEZ: Voting is now open for feasibility of Measure 0576. Option Number 1, high, Option Number 2, moderate, Option Number 3, low, and Option Number 4, insufficient.

All votes are in. And voting is now closed. The voting for feasibility of Measure 0576, 30 percent voted high, six individual votes, 60 percent voted moderate, 12 individual votes, ten percent voted low, two individual votes, and zero percent voted for insufficient. So for the feasibility of Measure 0576, this passes the criteria.

CO-CHAIR BRISS: So that moves us to usability and use. It's used in a variety of places. Anybody have additional comments that they'd like to make?

(No audible response)

MS. QUINNONEZ: Voting is now open for usability and use of Measure 0576. Option Number 1, high, Option Number 2, moderate, Option Number 3, low, Option Number 4, insufficient information.

Voting is now closed. For usability

and use of Measure 0576, 32 percent voted high, six individual votes, 53 percent voted moderate, ten individual votes, 16 percent voted low, three individual votes, and zero percent voted for insufficient information. For usability and use of Measure 0576, this passes this criteria.

CO-CHAIR BRISS: So with that, overall suitability. Anybody want to make closing arguments before we do the final vote? Hearing none -- I'm sorry, Rhonda.

MEMBER ROBINSON BEALE: We're voting on this based on how it is right now, not the planned changes that NCQA will be making, is that correct?

CO-CHAIR BRISS: Yes.

DR. BURSTIN: Although they did indicate several changes they were willing to make by annual update, I think, as you heard. So I think, you know, you could factor that in. But we are voting on the measure as it is now.

CO-CHAIR BRISS: Yes. We're always voting on the measure as it stands. So with that, yes or no?

MS. QUINNONEZ: If there are no more comments, we'll be voting on overall suitability for endorsement of Measure 0576. Option Number, yes, Option Number 2 is no.

Voting is now closed. For the overall suitability for endorsement of Measure 0576, 80 percent voted yes, 16 individual votes, and 20 percent voted no, four individual votes for the overall suitability for endorsement. Recommendation, this passes.

CO-CHAIR BRISS: So thanks to NCQA for hanging with us as we discuss the rapidly evolving landscape of behavioral health integration and how you communicate with patients.

(Off microphone comments)

CO-CHAIR BRISS: Not today. It's too late to talk about measure harmonization today. We can table it until tomorrow. If you we need public comment --

DR. LUSTIG: We do need public --

CO-CHAIR BRISS: So really, would any of you long-suffering people in the back like to

make you public comments? Would anybody from the phone like to make a comment?

OPERATOR: If you'd like to make a public comment, please press star one. Press star one to make a public comment. And there are no public comments.

CO-CHAIR BRISS: So thanks to everybody for hanging with us. This was quite a day.

DR. LUSTIG: For the committee members that are coming to dinner with us, it's really just down the street. So for anyone who's joining the dinner, it's literally down the street from here, so five minute or less walk. It's called Siroc, S-I-R-O-C. And everyone should have a -- the committee members, you should have a menu and address in your folder, actually.

I promise you tomorrow will be easier. And we start, I think it's at 8:30. And if you guys beat us there, the reservation's under National Quality Forum.

(Whereupon, the above-entitled matter went off the record at 5:53 p.m.)

E

G

O

Q

S

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 Standing Committee

Before: NQF

Date: 02-28-17

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R. Gross

Court Reporter

NATIONAL QUALITY FORUM

+ + + + +

BEHAVIORAL HEALTH STANDING COMMITTEE

+ + + + +

WEDNESDAY
MARCH 1, 2017

+ + + + +

The Standing Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:40 a.m., Peter Briss and Harold Pincus, Co-Chairs, presiding.

PRESENT:

PETER BRISS, MD, MPH, Co-Chair
HAROLD PINCUS, MD, Co-Chair
MADY CHALK, PhD, MSW, Treatment Research Institute
SHANE COLEMAN, MD, MPH, Behavioral Health Division Southcentral Foundation
DAVID EINZIG, MD, Children's Hospital and Clinics of Minnesota
CHARLIE GROSS, PhD, Anthem, Inc.
CONSTANCE HORGAN, ScD, Brandeis University
LISA JENSEN, DNP, APRN, Veteran's Health Administration
DOLORES (DODI) KELLEHER, MS, DMH, D Kelleher Consulting
KRAIG KNUDSEN, PhD, Ohio Department of Mental Health and Addiction Services
MICHAEL LARDIERI, LCSW, Northwell Health
TAMI MARK, PhD, MBA, Truven Health Analytics
RAQUEL MAZON JEFFERS, MPH, MIA, The Nicholson Foundation
BERNADETTE MELNYK, PhD, RN, CPNP/PMHNP, FAANP, FNAP, FAAN, Ohio State University*

BROOKE PARISH, MD, Blue Cross Blue Shield of New Mexico
DAVID PATING, MD, Kaiser Permanente
VANITA PINDOLIA, PharmD, Henry Ford Health System (HFHS)/Health Alliance Plan (HAP)
RHONDA ROBINSON BEALE, MD, Blue Cross of Idaho
LISA SHEA, MD, DFAPA, Butler Hospital, Care New England Health System
ANDREW SPERLING, JD, National Alliance on Mental Illness
MICHAEL TRANGLE, MD, HealthPartners Medical Group
BONNIE ZIMA, MD, MPH, UCLA Semel Institute for Neuroscience and Human Behavior
LESLIE ZUN, MD, MBA, Sinai Health System

NQF STAFF:

KAREN JOHNSON, Senior Director
TRACY LUSTIG, DPM, MPH, Senior Director
ELISA MUNTHALI, MPH, Vice President, Quality Measurement
ERIN O'ROURKE, Senior Director
DESMIRRA QUINNONEZ, Project Analyst
KIRSTEN REED, Project Manager

ALSO PRESENT:

PAUL CLEARY, PhD, Yale School of Public Health*
SOEREN MATTKE, Dsc, MPH, MD, RAND Corporation
ELIZABETH SLOSS, PhD, RAND Corporation
KATE WATKINS, MD, MSHS, RAND Corporation*

*** present by teleconference**

A-G-E-N-D-A

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

8:40 a.m.

CO-CHAIR BRISS: So maybe I will start. Welcome back. Thank you all for coming back after all the punishment yesterday. We got through a lot of measures yesterday. So we finished ten of them, several of them hard, and approved seven out of ten and still got out in time for dinner. So congratulations.

Today looks like an easier day. We will try to successfully get through our three today and have some additional time to discuss. It is kind of -- we should have some time to discuss what the portfolio looks like as a whole in the afternoon today. So, that should be kind of an interesting discussion that we don't always get to have. So, I have no further preamble.

Harold?

CO-CHAIR PINCUS: Just I think it is important to emphasize that discussion because part of our role is getting kind of a guidance to the field about what sort of measures are needed,

how have measures evolved, what are the sort of expectations for measures that get submitted in the future. And I think certainly that discussion will be informed by our discussions yesterday in terms of what were the criteria we really thought about in terms of looking at the measures that were approved and those that were not. And so we are the evolving sort of state-of-the-art of measurement in this area is going.

So, start thinking about that and sort of jotting down ideas. And we will be able to look at some of the alignment issues, like especially for example tobacco, where we had some discussions yesterday about the varying types of measures that are applied in different context and how to sort of make them more aligned.

Any questions? Mike.

MEMBER LARDIERI: Could we talk or maybe have someone explain a little better the economics of measures? Because a lot the developers, some of them do them because like CMS they want to do them because they need to do them

to move things forward but some of the other measure developers, don't they get paid for this? What I read in going through some of the materials and I knew it but it never really dawned on me that when they develop a measure, they are going to push it to an EHR, that EHR vendor needs to pay them a license in order to access it. So, I am wondering if that has something to do with why we have so many different measures that are like with tobacco. They are doing the same thing, different developers doing a little different something but then everybody pays downstream. And I don't know if I am right or wrong on that but I would like to know the economics of it a little better.

CO-CHAIR PINCUS: Okay, so you are talking about both the economics of how measures get developed and the economics of how measures get implemented?

MEMBER LARDIERI: And how they get -- who pays for that. Because as a provider, I'm paying for it at some level. If a measure

developer is charging an EHR company a licensing fee, that licensing fee is passed down to me. So that area of how it works.

CO-CHAIR PINCUS: Okay. Mady.

MEMBER CHALK: Yes, the other part of that, which is bedeviling, some organizations that are not sending measures forward, despite our best efforts, is the whole cost of stewarding a measure. It is quite costly but essential. Otherwise, measures can't be brought forward.

But there are several organizations that I am talking to about bringing measures forward that simply can't do it. And there is no obvious source of funds for that and the cost is substantial.

So maybe when you --

CO-CHAIR PINCUS: No, I think that is an important issue in terms of as we identify gaps, how exactly do those gaps get filled in a practical way.

MEMBER CHALK: I was hoping that at some point in our open time of discussion we could

begin talking about the whole question of proximal outcome measures and what happened with the tobacco measure yesterday. That is a proximal outcome measure but couldn't get through for specification issues.

There are lots of instruments that are being used standardly in the field, both in mental health and addictions treatment and that could yield proximal outcomes. The ASI -- well PDQ-9 we know about and the ASAM criteria, those are not coming forward and there is a reason -- because they don't think they can get through. And I am hoping we can at least talk about what it is going to take to get those to change that view, I guess, of the NQF process and think through how we can encourage those kinds of measures.

MEMBER GROSS: The second and third Mike and Mandy's point and I think it is another facet of the same issue, maybe at some point we can discuss what is NQF's role, what position do they want to take relative to being the thought leader

with regard to multiple measures and narrowing around proximal outcomes and also encouraging potentially new ones because I think there is a real opportunity. There are so many measures out there that are costly. The costs go somewhere. So people are either paying multiple times for similar measures or choose not to do any of them. So I think there is an opportunity for some organization to be the leader in this area. And I wonder NQF's role in that. So, maybe we can have a fruitful discussion about that sort of thing.

CO-CHAIR PINCUS: So, I don't know. Is Shantanu or Helen going to be here later?

MS. MUNTHALI: Helen will not. I'm not sure about Shantanu. But with regards to what Mady was saying and I think everyone else around the table, measure development is very costly. So by the time measures come to us, it really is too late. There has been quite a bit of investment that has been made in putting these measures together. So our focus and priority has

been on how we can affect the upstream.

What you will be doing in terms of identifying the right gaps, the right measures will help us to communicate that to developers through our Measure Incubator and developers that come to us.

But then, as Mady said, our process is another challenge. We do have a scientific rigorous process. We are hoping, again, that we can get to measure developers through technical assistance. This is something that, in the last year or two, we have been pushing to hear from measure developers earlier than we have because we do know, when they are around the table, they haven't really gotten the input that they need to be more successful through our process. But I think we can talk about it a little further when we talk about gaps.

CO-CHAIR PINCUS: Yes, I think it would be good to give a little bit of an update to people about the Measure Incubator and how that works or how it has been working and what the

plans are for that.

Mike?

MEMBER TRANGLE: Two years ago as part of his plenary at IHI, Institute for Healthcare Improvement, it was two or maybe three years ago, Don Berwick's keynote speech talked about how there were too many measures and how we need to sort of figure out how to simplify and reduce the number.

And I think if we are thinking about the role of NQF, I don't see an equal emphasis on how do we synthesize three measures down into one. How do we sort of like -- you know, it is all about new ones coming up and then when they happen to come up again, you talk about them and we sort, in a half-assed -- excuse my language -- fashion, without that much time or real scrutiny or effort to talk about harmony. But I don't think we really harmonize. We just talk about it.

We probably need to have equal effort and equal pre-work about synthesizing measures as

we do in creating new ones.

CO-CHAIR PINCUS: I'm a little bit worried about having the discussion we are going to have later now. Because I think when we look at what is -- and I think the discussion of the three, the two and a half measures this morning will also I think add to the discussion later on because they bring up exactly some of the issues that you guys are talking about.

CO-CHAIR BRISS: And clearly, we will have to think about the agenda because it is clear that nobody is going to have anything to say about the portfolio overall. It is kind of disappointing but we will try to do the best we can this afternoon.

CO-CHAIR PINCUS: But any -- Rhonda and Tami.

MEMBER ROBINSON BEALE: I shouldn't bite into a bagel.

I just want to bring up an issue I think is parallel to what has just been discussed and that is I know that there is a lot of easier

measure development when we are looking at structure and process. Outcomes is a whole different story.

In behavioral health, probably even more so than other fields, there is a tremendous need for outcomes and being able to measure the outcomes.

I know there has been attempts in the past to bring forward a measure that is a technique that is used to determine whether or not treatment has been effective. And that is looking at the effect size. It is looking at the amount of change that has occurred over a period of time. It doesn't fit into a nice denominator and numerator type format but it has a lot of validity in terms of really being able to determine whether or not there has been an effect based on the intervention and it is a methodology that can be used across many different types of outcome measures.

And I guess when Mady was talking about the fear of coming back again, when I talk to

those who are in this type of -- doing this type of work, they don't even think about the NQF because they feel as though they are not, this format is not going to be accepted. And I am wondering if times have changed now because I think the last time was about six years ago, whether or not thinking has changed or whether or not there is more thought around this type of approach.

CO-CHAIR PINCUS: So, let's move ahead with the discussion of -- now it is not actually -- we are not considering this as for endorsement but as something that we want to have some discussion about. And, hopefully, it will also pertain to our discussion later because this is a perception of care, a kind of an outcomes sort of measure that has come up for maintenance. And also one of the issues is is that there has been little attention to this in terms of new information about its use. It is a much older measure.

Is Paul Cleary on the phone?

Are we connected to the phone? How do you call the operator?

MS. REED: Operator, is Paul's line currently open? Because he is saying that he can hear us but we are not able to hear him.

OPERATOR: Yes, his line is open.

DR. CLEARY: Hello? Operator? Operator, can you hear me?

OPERATOR: Yes.

MS. REED: We can hear you.

DR. CLEARY: Can you tell the committee that you can hear me?

CO-CHAIR PINCUS: Yes. Hi, Paul, this is Harold Pincus. We can hear you.

DR. CLEARY: Oh, okay, great. Hi, Harold. Hi, Peter. Hi, committee members.

CO-CHAIR PINCUS: Good. So we are just starting a discussion about the ECHO measure. And as I think Helen spoke with you earlier, and said that because there is so little further information on it, that it is difficult for us to formally consider this for endorsement but it is

something that we want -- because of the importance of perceptions of patients is so important, we wanted to get an understanding of where this measure stands in terms of its updating, in terms of its use, in terms of what additional data there is with regard to both this measure specifically and how it fits in with any measures that might be in development.

DR. CLEARY: Okay, do you want a three-minute summary and then I will open it up to questions?

CO-CHAIR PINCUS: Sure.

DR. CLEARY: Okay, so the first point I will make, this may be obvious to everyone, but perhaps not, it is part of what call the CAHPS family. And CAHPS stands for Consumer Assessments of Healthcare Providers and Systems, so a group of surveys that are stewarded by AHRQ to basically assess patients' experiences with care.

As someone alluded, I think Peter mentioned quite a while ago, it was back in the

late '90s, actually, Ken Manderscheid as SAMHSA came to us and was very interested in expanding to behavior health and there was a tremendous amount of interest in the field in general. And at the time, a survey called MHSIP, which actually is still used, was the most widely used one. There was some questions about it whether it was the best survey and/or could we have a measure that was part of the CAHPS family. A detailed -- there was quite an elaborate process with different professional and advocacy groups to try and take the best of the MHSIP and some current CAHPS surveys. And we came up with ECHO to be sort of politically neutral but it is really a CAHPS survey.

And regarding the recent discussion I just heard, I consider CAHPS measures process measures; that is, what did or did not happen when you saw your behavioral health specialist.

During the development process, there was a tremendous amount of interest in and really quite adamant support for some assessment of

outcomes. Now, there is no delusion that this is an outcomes measure but there are some measures in there about perception of improvement. I was not in favor of those because I don't think this kind of measure is the best way. You know there are symptom scales for depression and anxiety and so on and those are probably better quote, unquote outcome measures but we did include some measures that have to do with perception of benefits from treatment. And I was finally convinced that those were reasonable because in behavioral health, ones perception of whether one was helped had some validity. So that is why it is a little bit of a hybrid and why it has the new name ECHO.

So, it is not really an outcomes measure but it is what I would call a process measure. It is part of the CAHPS family.

One, talking about stewardship, so the CAHPS Consortium Stewards Survey, all CAHPS surveys, we spent an enormous amount of time revising, updating, testing and so on. And as

people have alluded, there hadn't been --

Oh, and the other thing that has been done by AHRQ is there is a CAHPS database. So people can voluntarily submit health plan CAHPS or group clinician CAHPS and of course, CMS has HCAHPS nationally. So, in those surveys, even though the CAHPS Consortium is not a data collector, we are an instrument development team -- and by the way, all the instruments are -- everything about the instruments is in the public domain. We don't get any financial remuneration when they are used, of course.

So with those instruments, it is relatively easy to pull data from the CAHPS database or from the CMS data and to do the types of analysis I think you would like to have seen this morning, such as the variation across providers or the average scores, or increases in scores of these particular instruments. We don't really have that just because it was never part of the database. What we did for the application is we do get routinely downloads of the

instrument and request for information about the instrument. So that is really about the best I could provide to you in terms of use. And I just --

CO-CHAIR PINCUS: Paul, can you say who the we is, a little bit more about the we?

DR. CLEARY: I'm trying to remember when I said we. I am a principle investigator of one of two teams which are cooperative agreements with AHRQ Ron Hays directs. Ron Hays, Marc Elliott, and I and Susan Edgman-Levitan direct another. So I am at Yale. Ron and Marc are at UCLA. We are the sort of CAHPS Development Teams.

And when I say the consortium, that includes AHRQ. Caren Ginsberg is our Project Officer and we do a lot of work with CMS. So, for example, a lot of the HCAHPS development was developed with CMS. We are developing ACO CAHPS and so on.

Is that what you were asking?

CO-CHAIR PINCUS: Yes, just to get a

sense of who owns the ECHO.

DR. CLEARY: Oh, I see. Well actually AHRQ owns -- AHRQ has the trademark to all the CAHPS instruments, to the word CAHPS. And at the time when Chuck Darby was still alive, we actually got the word ECHO trademarked.

Let me just fast forward because one of your questions in the introduction was what are we doing. There has actually been a resurgence of interest. And we submitted for maintenance, just because our maintenance cycle is up, to be quite honest. But I would say over the last year or so we have been getting increasing interest in using the instrument. We have done some cognitive testing, some focus groups. We are about to field test a subset of items in the State of Connecticut. And we actually, if we get the funding, are proposing a major revamping or updating of the ECHO and we will probably called it Mental Health CAHPS.

And in the process of doing it, I cited a number of studies. So when we developed it, we

did some fairly large studies with different partners, both providers and professional organizations, and then haven't done much of that in the past. For example, Massachusetts General Hospital is going to be using an updated version that we developed on a PCORI project but I just don't have much new information now. So that is like sort of the 30,000-foot overview. But let me stop and see what you would like more information about or how I can be helpful.

CO-CHAIR BRISS: Paul, this is Peter. Can you give us a little bit more about you just said that there is increasing interest in using. And so can you give us a little bit more detail about what kinds of people or organizations are showing the interest and how they would like to make use of the data or results?

CO-CHAIR PINCUS: And also, when you said you are beginning to do some additional studies, is that with funding from CMS or from AHRQ? What is their intent in terms of this will impact on ultimately its application?

DR. CLEARY: Okay, so two questions. So, as I mentioned, the PCORI Group at Mass General Hospital is doing a large study. I believe it is on the treatment of depression. They were interested on the perception of care.

When we developed the instrument, many of the users were Medicaid state programs. So for example, the State of New York was using it for a while and some other states have used it. Some insurers are interested in using it. We just had a call this week with three of the major health insurers, Anthem, United, and ConnectiCare in Connecticut of possibly doing a survey. The State of Connecticut has approached me reviewing a state innovation model evaluation for CMMI there. One of the priorities of the advisory groups is trying to get a pulse of how well we are doing with mental health treatment.

So for example, we are fielding a statewide CAHPS survey but we are including a subset of ECHO items in that, starting to get a sense of that. And I am going to propose with

either the insurers or the State Medicaid offices a large survey.

We also are doing a big study with community health centers, 12 community health centers in Connecticut and they have expressed an interest in assessing the behavioral health services. They have a very integrated -- they are an FHQC. They have a very integrated model. They are very interested in assessing the quality of their care.

Regarding the second question, the AHRQ. So I have a cooperative agreement with AHRQ. We do a whole range of things related to CAHPS. And one of those things, as I said, we have started doing are like focus groups and cognitive -- we have done I think like four rounds of cognitive testing. And that is paid for by the cooperative agreement. And on this year's budget we have proposed a major revision and field testing of what we think will probably be called the Mental Health CAHPS.

CO-CHAIR PINCUS: So a number of people

have sort of raised their cars around the table. So, Vanita?

CO-CHAIR BRISS: So, again, for this one, we are not trying to approve a measure the way we were doing yesterday or the way we will do for the last two today. We are trying to have a conversation and talk about trying to get a better sense of some of the questions that some of us couldn't answer easily through the application materials, like how is the measure being used. And we are trying to --

DR. CLEARY: I'm sorry to interrupt. As you go -- during the discussion, I am eager to hear what kinds of information would be most useful to you in the future for making a decision. So, we have certain kinds of information we routinely collect but you may have a different set of priorities or different interests. So, I will be eager to hear that.

CO-CHAIR BRISS: And so this is a little bit different kind of a conversation.

Paul, maybe I will take a chair's

prerogative and ask you a question to start with. As I hear you discuss this, this is a bit of a -- it sounds a bit less like a standard quality measure and it sounds more like, in some ways, a public health surveillance system or a survey instrument or a more science kind of project.

So can you help me with is this really a quality measure? And does NQF endorsement help you use this instrument in the way you are trying to use it? This is a bit of a different kind of thing, at least as it has been teed up I think. I would love to hear you comment on whether you really see this as a quality measure and not a more surveillance or science project and would endorsement or re-endorsement help you; and if so, why?

DR. CLEARY: Sure. So it is probably my fault for talking like an instrument wonk. I do that kind of stuff. But I would absolutely say it is a key quality measure.

Let me just start. The CAHPS instruments, we consider patient-centered care to

be a key component of quality and one of the best of ways of assessing that is by listening to patient experiences.

So for example, HCAHPS is now used nationally to determine a percentage of reimbursement for basically all hospitals in the country. Health Plan CAHPS is used by NCQA as a quality measure in every accredited health plan. I think like 30 Medicaid programs use CAHPS surveys to assess quality of care.

We are now using CAHPS surveys in Connecticut to determine payment on shared savings programs as part of the main insured. So they look at patients' experience, how they vary and we have a contract with CMS on both hospitals, ACOs, now and the new MIPS surveys. All those are quality measures that CMS is actually using to determine reimbursement on their quality metrics.

So I talked about the development process just because that is what we do and the way to it being used as a quality measure but

really the only reason for developing them is to assess and facilitate improvement of care quality.

CO-CHAIR PINCUS: I think that one of the things that makes this somewhat unique is I have, and a number of other people, you know and the integral study that you are talking about at Mass General that use the measure in the context of research studies or use elements of the measure or certain items in the measure. So it gets a little bit murky in terms of its overall use.

DR. CLEARY: Okay, that is a good point. That particular study, I agree with your characterization. And our interest in doing it was A) to be helpful; and B) we thought that we would get a lot of -- we used a variation. We have made some modifications. It was an opportunity to test psychometric properties and so on. So, in terms of that particular study, you are right. That would be, in their opinion, maybe a surveillance study; in my opinion, part

of the development process.

But would NQF be necessary or helpful? This comes up a lot with our instruments but, as you know, many payers or CMS require NQF endorsement to use a survey in a program. So we want to use this eventually in our SIM program. I can't answer you exactly but I am pretty sure that endorsement would be very, very helpful when we go to the State of Connecticut. I am now leading the evaluation of their SIM project to say we want to use this statewide to evaluate the quality of care in Connecticut.

CO-CHAIR PINCUS: Okay, so why don't we go around? So, Vanita and then we will come back. Vanita, and then Tami, and then we will go up the other side.

MEMBER PINDOLIA: Good morning. Thank you for sharing a little insight about how the outcome surveys were developed with feedback from providers and others.

Could you expand on the patient focus groups that are included for the survey

development?

DR. CLEARY: Sure. Originally, I don't have an exact count, I'm guessing we did 10 or 20 focus groups. And the ones we do now, there is several iterations of information that we get directly from consumers. One is, especially during the original development and we would redo this if approving to upgrade the ECHO is confirmed, we talked to patients who have had behavioral health services and go through a semi-structured protocol to understand how they perceive the quality of care, what they experienced, what they think about, what they mean about quality. We also do extensive reviews of the literature, what we consider appropriate communication, et cetera, et cetera.

We did a Federal Register request for all available instruments and so on. We often, then, develop a preliminary instrument and we will have focus groups where we combine what I would call cognitive interviewing with focus groups. So you will sit down with a group of

half a dozen people, have them go through the survey. And there is two levels at which you try and understand the instrument. One is, what was their understanding of Question X? So, I will give you one example. One big issue that advocates were very concerned about was a feeling of safety. Well, it turns out that is a very hard concept to get at, what you really want to know about in terms of safety. So you have a focus what did safety mean to you. How did this term interpret it? You know and you ask it two or three different ways just to see if they understood the questions, how they interpret the questions, and what their responses meant.

And we go through iterations with different subsets of patients. They tend to be homogeneous. We might have a -- we will for a CAHPS instrument always do focus groups with people whose first language is Spanish, maybe a group of people on Medicaid, people with private pay and so on.

CO-CHAIR PINCUS: Tami.

MEMBER MARK: This is really informative. Thank you very much, Paul.

Can you talk about the extent to which the current ECHO is relevant to the substance use disorder population? And you mentioned that you might be renaming it the Mental Health CAHPS. So does that mean the future CAHPS is not going to apply particularly to patients with substance use disorders?

DR. CLEARY: No, we originally used behavioral health for the reason a lot of use the term behavioral health. We would probably change it because when we have done preliminary focus groups and cognitive interviewing, virtually not a single patient identified with the term behavioral health. They just didn't know what we were talking about.

So in the survey there we try and frame it very, very broadly and substance abuse absolutely would be an emphasis, especially at the time it was a key component and with current issues we absolutely will focus on that and try

and understand how we can get at those experiences.

The renaming is purely a -- people just didn't understand the term.

CO-CHAIR PINCUS: But I guess the question is so it would be designed to be applicable to groups with substance abuse?

DR. CLEARY: Yes.

MEMBER MARK: Thanks.

CO-CHAIR PINCUS: Let's go up the other side.

MEMBER KNUDSEN: Hi.

DR. CLEARY: And by the way, just when we originally developed with SAMHSA, of course, there was tremendous interest through SAMHSA for substance abuse or a range of treatments now.

MEMBER KNUDSEN: Okay. This actually follows that. You had said that SAMHSA had asked you or requested you to basically help develop this but they didn't endorse it and they continued to use the MHSIP. So, I was just wondering --

DR. CLEARY: No, it wasn't that they didn't endorse it. Actually, Ken was very -- it wasn't -- I wouldn't say it was official. Ken came to us and said -- you know MHSIP actually refers to a broader program. Within that, there is a MHSIP survey. And it was a long time ago. I don't remember the specific discussions but they were very interesting. We went through a very, very elaborate process and they didn't endorse one or the other. We did adopt ECHO.

My only comment was that a number of states continued to use MHSIP. They had been using MHSIP. They had historic data on MHSIP.

By the way, the CAHPS Consortium has -- we try and develop standardized surveys with protocols, with analytic programs, with supports and report formats, put it in the public domain. Whether people adopt it or not is up to them. AHRQ says nothing to any -- I mean CMS has done this in many instances. We just try and make available an instrument that people will find useful for quality assessment and improvement.

CO-CHAIR PINCUS: Okay, Mike.

MEMBER TRANGLE: I want to share a perspective that is really sort of a delivery system-based perspective. From a system that is regional, pretty large, very active, and kind of quality and patient satisfaction work, really sort of pursuing the triple aim, I would say it seems, for like the last five or ten years, the kind of questions we have are should we just have our CD, you know substance use disorder patients and mental health patients in with the general survey. If that is true, is there any way we can get a subset to benchmark ourselves against a like cohort, whether it is just mental health or combined mental health and CD?

Do we do the behavioral health? And even then, when we have done it, could we get any benchmarking to figure out where we are at and, ideally, find out places that are doing better that we could talk to them and learn from them?

We actually, our vendor was NRC Picker. About two or three years ago, we forced them and

they did sort of like a teleconference to see who is doing well and who is not and how to learn. And it was such a polyglot group. Some of them were inpatient CDs, some of them were outpatient CD. Some were large urban, like ours, 100-bed psych hospital, 100 psych beds in a bigger hospital. Others were like an eight-bed unit out in the rural hinterlands. You know, I am not even -- we never hear of in the cities.

And it just felt like walking in molasses or something to try and sort of just get down to reality and do something useful with the data that we know is important. And in some sense my perspective is less esoteric; is it viewed under quality or is it separate but equal, equally important and how it fits into NQF?

Are you going to sort of start answering some of the questions that my system and I think other systems like that are trying to get answered? And I could be pretty agnostic if there is a large enough end within the general HCAHPS for inpatient or ambulatory CAHPS and we

could compare ourselves to behavioral health patients, I would be okay with that versus a separate survey but it would be really nice to have one of them sooner versus later.

DR. CLEARY: Boy, if I could answer that. That's a great question. I will give you a quick background and then what our current thinking is.

The quick background, back in 2000 or so, we -- the ECHO actually has two versions. There is a health plan version and then there is a Behavioral Healthcare Organization version because carve outs were very distinct and consumers identified with them and so on. And the ECHO, unlike CAHPS surveys, CAHPS are very anyone who has received care in the last six months, ECHO is designed and used by people who could identify behavioral healthcare users.

So, for example, with the State of New York, we identified people who had a diagnosis, had a psychotropic medication, had -- there was three or four condition and we used their record

systems to identify a cohort of people who had received behavior health services. So, that was a distinct survey and their approach was how are patients in Plan A, B, and C receiving their behavioral health services.

The focus groups we have been doing over the last year or so indicated -- I think you have characterized the situation well. We think what is likely to emerge is a subset of questions for a general population about access to a behavioral health services and a survey for people who are known to have received behavioral health services to assess quality of that care.

So for example, the State of Connecticut is going to do a CG CAHPS survey statewide as part of their state innovation model evaluation. They wanted behavioral health items and we included I believe six items. We did some cognitive testing and some initial, many field tests and so on. And they are basically you know did you have the need for behavioral health services; were you able to get some sort of

access to care items? And we are going to test those statewide.

And what many people have been telling us, again, this will be part of the development process, is it would be good to have something for a general cohort of patients, many of whom everyone in this room knows have behavioral health problems to get a sort of 20,000-foot assessment of need and access and then a more detailed instrument for a drill down.

One thing I will just mention, these are all issues that are very difficult when you are developing things, one of the ways in which the ECHO is very different from some of the other CAHPS surveys, so for example, CG CAHPS, the Clinician Group CAHPS, it focuses on a specific provider group. That is very important because attribution is part of the quality. When NCQA gives a score to a health plan, they want to know that that care is attributable to that health plan.

As people in this room, behavioral

health services are provided by a wide range of providers. There is also often very complicated comorbidities. And so it is a much broader thinking about the range of services that one got.

And so that is, in terms of your question, a balancing act is knowing what is the overall quality of care that the person received as opposed to what is attributable to your system and what do you want to be accountable for and/or approve.

So the short version of that is I think we are probably going to have -- we are thinking right now our goal is to develop a set of items, you know five or ten items that could be a supplemental item set for health plan CAHPS or CG CAHPS or even HCAHPS. There is a lot of interest in inpatient treatment and that is a separate ball of wax and a survey that would be more detailed for people who have received behavioral health services.

CO-CHAIR PINCUS: Okay, Connie. You're

microphone is not on.

DR. CLEARY: I can't hear her speaking.

MEMBER HORGAN: Oh, sorry. Hello, Paul.

DR. CLEARY: Hi.

MEMBER HORGAN: My questions have already been asked by Tami and Mike. So, thank you.

CO-CHAIR PINCUS: So, I guess I was next in the queue. And I actually have three comments and a question.

One is similar to Mike's in terms of thinking about the particular target uses of this. And I have been familiar with its use in terms of evaluating health plans and behavioral health plans but to think about if you are going to the trouble of developing these items, and so to think about their adaptation for other settings, in particular, sort of outpatient specialty behavioral health settings. I would be interested to see whether it can be generalized between substance abuse and mental health and whether there might need to be some specific

items that relate specifically to substance abuse, potentially.

And in particular, something that came up and I know that this was something that came up on the MAP side of things for the Inpatient Psychiatric Facility Reporting Program, that there was an early effort by CMS to put in place a survey for inpatients that was going to be part of that program. And that got rejected by the MAP because it was an adaptation of a survey that was developed for adolescent inpatient settings that happened to have been an endorsed measure.

And so something, if there is going to be this program on inpatient reporting, having a CAHPS ECHO kind of thing that is aimed at inpatient would also be useful for that program and to think about that.

Number two, when I was at Pittsburgh, our health plan actually was doing ECHO surveys and they found that there was actually a substantive difference between different ways in which it was administered so that whether it was

administered by mail or by phone, that there was sort of a different level and type of response.

And so it would be useful now, especially also, the potential to administer it electronically by email or through an open website. So, I would hope that that would be part of the analysis and work that you are doing.

Number three is the original ECHO was out before really the recovery movement was in full swing. And to look at whether there are certain aspects of a recovery orientation that could augment the work that you are doing in terms of the kinds of items that you might be selecting.

And then the question I had is why has it taken so long for the CAHPS group to come back to looking at behavioral health or mental health?

DR. CLEARY: Okay, quick answers. First of all, I hadn't -- I will admit I didn't even know about the inpatient measure until I saw it referred to in some document. So, I wasn't part of that. I wasn't aware of that. That gets

raised periodically with us. I will definitely raise it. That is a big deal, so to speak. So my guess is it is not going to be done this iteration because it just takes a lot of effort but I will put that on the agenda.

We always do testing with mail versus phone and now spending a lot of time, one of the reasons for going with MGH because it is a portal type data collection, we do randomized studies of mail versus phone to assess mode effects. In some of the national surveys, we have mode adjustment factors and so on. So that is definitely part of any CAHPS development protocol.

Thanks for your comment about the recovery orientation. It is a good point and that will be useful as we go forward.

Why is taking so long? It is just we have a lot of activities that have been going on. I mean the good news is that CAHPS is very widely used. Now the bad news is it is very widely used. So we have spent an incredibly intensive

effort focusing on like Hospital CAHPS, Clinician Group CAHPS, working on MIPS and ACO and so on. And it is just, maybe like in many cases, it gets less attention than many of these other areas.

We have sort of know that it is -- and we just started a general updating process I would say a year or two ago trying to reconcile all the various instruments because minor modifications develop over time.

And by the way, back to the -- there was an earlier discussion what roles should NQF play. We are very worried about the proliferation of CAHPS instruments and so we have spent a lot of time trying to standardize cores, make them as compatible as possible. When people come in with a new survey we say now if you are going to do a cancer survey, use the same communication composites, et cetera, et cetera.

The big issue in the field that we see is the overlapping and multiple mandates. So what we are scratching our heads just nationally how to do is to sort of coordination of these

various mandates.

So for example, in the State of Connecticut you may get a survey as a Medicare beneficiary because you are in an ACO, because you are in an NCQA accredited health plan, because you get HCAHPS. Most people don't but it is these overlapping mandates. And I know that is not NQF's problem but that is the bigger issue in our mind.

And so I guess the short answer is we just haven't gotten to it. No good answer.

CO-CHAIR PINCUS: Peter.

CO-CHAIR BRISS: Paul, this has been a really helpful discussion. I have a comment and a couple of pieces of advice.

And the comment is I think the world has changed some since this was first developed and we are kind of in a world of more primary care integration. So were I you, I might think a little about whether you want to limit this kind of measure just to people that are seen in specialized behavioral health settings. I think

it might deserve a broader patient population. It is something for you to think about but it looks like there at least a few head nods around the table when I say that out loud.

And then a couple of pieces of advice. You know the NQF process is kind of very specialized in Byzantine. And there are all kinds of things that came out in your teeing up of the measure this morning that I didn't get clearly, at least on my read of the materials. So I think if you want this to be NQF-endorsed, I think you could get it endorsed. I think you would benefit from some help from somebody who does this a lot. So, whether that is NQF staff to try to help you tee it up to fit the kind of pro-crusty bed of NQF jurisprudence or something like that I think, or somebody else that has done this a lot, I think that would really help.

And then a couple of things that as you are doing that that I think might be beefed up. This is a fairly complicated measure with a lot of moving parts. And so in the initial framing,

I think you might do more about how these kind of patient-reported outcome kind of things relate to structures and process of care, I think you might do better about the kind of logic modeling that would make it easier to follow the measure.

And especially when you are bringing measures back, there is an increasing focus here on kind of usability and use. You know who is using it and for what and why. And so the kind of stuff you talked about this morning, so people are using it for state SIM projects and so on and so forth didn't come out to me clearly in the materials. And that, in particular, might be beefed up more than it is.

Thank you for talking to us this morning.

CO-CHAIR PINCUS: I like the NQF jurisprudence. That is sort of interesting. I use the term quality measurement industrial complex.

Rhonda?

MEMBER ROBINSON BEALE: Okay, thank you,

Paul. I have some questions for clarification and also to get a better understanding.

So I didn't read through all the materials but I did read through and was a little bit confused by the numerator/denominator that you offered. I was wondering if you could explain how you were conceptualizing that using this ECHO survey.

DR. CLEARY: Sure. Basically, as I mentioned, the denominator was conceived as people who have received behavioral health services. And by the way, this was applicable to primary care, as well as specialty care. And that is defined based on either the -- usually record information has been used in the past but you could ask screening questions of a general population.

So you have let's say a thousand people who have used behavioral health services in the last six months and then you say did someone tell you about the options for treatment in a way that you could understand. And you calculate the

number of people who say always, sometimes, you know yes versus no. So 60 percent of the people -- I was just looking up one number in the application. Did you try and get a referral? In one system it was 67 percent of the patients were able to get a referral and in another part of the system it was 85 percent.

So the idea is the denominator is people who have used behavioral health services, what proportion were able to get an appropriate referral and then people who say yes is the numerator.

MEMBER ROBINSON BEALE: Okay, so you were talking about using individual elements within the survey itself.

The second question --

DR. CLEARY: Yes and then maybe this is too wonky but CAHPS surveys are really, the main goal is for assessment of providers or systems not -- they certainly can be used at the individual level. So you can monitor a patient and say they did or did not get a referral or

were black patients less likely to get a referral than white patients. We do a lot of those types of analyses but the main CAHPS Analysis Program, it is called our macro create scores at, let's say a primary care clinic, or a primary care group, or a behavioral healthcare organization. And then those numbers are rolled up to say you know the Yale Medical Group of the people who used behavioral health services, 70 percent said they could get a referral whereas as in community health centers, 90 percent said they could get a referral.

CO-CHAIR PINCUS: One clarification on your response to Rhonda. So, does that imply that for users of the survey they don't -- they can pick and choose which items they would include in the survey, since it is an item level?

DR. CLEARY: Well, I mean people can do whatever they want to do. I don't mean to be facetious but CAHPS feels very strongly because we know that there is order effects and context effects in the survey that to be called a CAHPS

survey, they should use the entire survey.

MEMBER ROBINSON BEALE: Okay, thank you. My last -- my next question had to do with whether or not there are benchmarks established for this. So that one, I think Michael that brought it up, if you really wanted to understand whether one system is functioning better than the other and to understand what is best practice, are there benchmarks established?

DR. CLEARY: So the short answer is no. Basically and, again, what the consortium does is makes available data. The CAHPS database accepts voluntary submissions. So you can go into the CAHPS database and look at in Medicaid programs what is the 90th percentile of plans where people said someone answered their question in a way they could understand. So you can look at the distribution of CAHPS scores. Or the State of New York did it in their entire Medicaid population and they could look at the distribution and look at -- one of the things we always calculate and is in some of the articles I

submitted is inter-unit reliability. One of the criterion we use for selecting questions is not sort of the individual psychometrics but because we like one to make assessments of providers or systems, inter-unit reliability.

So when we do it in Connecticut is are the different groups that are assessed, do certain items reliably differentiate from the groups and then they use whatever data they have to say what is our benchmark, if you will.

So an example is we do HCAHPS for CMS every year. We calculate regionally a whole variety of adjustments in waiting and so on but then calculate distributions of the HCAHPS scores and the Star Rating System, if you will, comes out of the assessments of the distributions in an individual hospital, vis-a-vis the regional distributions.

So there is no absolute -- you know you could argue an absolute -- you could say what is the best performing system in your -- best performing group in your system but we don't have

like an absolute everyone in the country should have an 87.

MEMBER ROBINSON BEALE: Okay, thank you.

My last question has to do with the section that is called perceived improvement and whether or not that has been tested at all or cross-referenced with actual measurement of outcome, where you are looking at the perception of patients in terms of perceived improvement against actual measurement of outcome.

DR. CLEARY: The short answer is no. In terms of general improvement, really the only -- for a lot of reasons it is obviously very hard to assess rigorously but since HCAHPS has done nationally for many years and Medicare CAHPS has done nationally for many years, we have done analyses of improvement in HCAHPS scores and Medicare Advantage and Fee-for-Service scores and have written a number of articles showing how those scores have improved. We think they improved more after public reporting of HCAHPS scores, et cetera, et cetera. So, we could show

that.

Whether it is related to outcomes, again, the answer, we have no idea in terms of mental health. The couple studies I have been able to do, it is a very hard question to answer but I did a study of all heart attack patients in New Hampshire and heart attack patients nationally in the VA and showed that -- I know this isn't behavioral health -- but showed that patients that had better patient-centered care, were their questions answered, et cetera, et cetera, et cetera, and the New Hampshire study actually had better outcomes a year later in terms of dyspnea and angina. And in the VA study, we actually showed a significant impact or relationship with mortality using a national data set from the VA. But we haven't even tried anything comparable in behavioral health.

As you might imagine, the design issues are very daunting.

MEMBER ROBINSON BEALE: Thank you.

CO-CHAIR PINCUS: Actually there may be,

just to put out it there, there may be some ways of actually doing some of that testing now that at least there are places that are doing sort of consistent measurement of PHQ-9 scores as part of actually an NQF-endorsed measure. And that might be sort of a place where you could actually test what Rhonda suggested.

DR. CLEARY: Yes, it's --

MEMBER ROBINSON BEALE: Well there is also another organization called ACORN, which does a lot of extensive data collection in terms of sets of providers who are using standardized tools during the treatment. And so you have beginning of treatment, and during treatment, and outcome and they actually are able to calculate effect size. So that could be a very -- and they have a large database. So it might be a really good partner to have them administer this in correlation --

DR. CLEARY: Is that A-C-O-R-N?

MEMBER ROBINSON BEALE: Yes, A-C-O-R-N, Jeb Brown.

DR. CLEARY: Okay, thanks. Yes, some of these things you know we have large EHRs and so on. So I think it is much more feasible than it used to be. That is a great idea. Thanks.

MEMBER ROBINSON BEALE: Okay, thank you.

CO-CHAIR PINCUS: Mady.

MEMBER CHALK: I'm going to try to be brief about this. Just because MHSIP was raised, when Ron Manderscheid and I, about 12 years ago, decided we were fed up with MHSIP and a variety of other things, Connie will remember this, under the auspices of the Washington Circle, we created and worked on a perception of care survey. We weren't happy with the CAHPS at the time. Ron wasn't. I wasn't. I was in the federal government. He had money. I had money. And we used it to fund this effort to develop what became known as the Modular Survey. Adult, adolescent, mental health, substance abuse disorders. Each one was about 10 or 12 items, pretty short.

We got it to be web-based, et cetera, et

cetera, and tested in a couple of places. And then you know it needed further funding. I left. Ron left. And there it sat. I have still got the disks sitting with me. It's ready to be -- I will give them to you.

The point being it was the kind of perception of care survey that allowed whatever population were working with, whether it was at the treatment program level or in aggregate to be administered in pieces without having to look through an entire survey and then try to subset it. All the items had had IRT testing. I mean it was -- it is, continues to be the kind of perception of care survey I think we need in behavioral health.

I don't know if anything will ever happen with it but we may offer it to Paul, if he wants it and wants to see how the items subset with the items that are in ECHO or somebody else could pick it up.

DR. CLEARY: Thanks. It is coming back. If someone could send me a reference so I could

just remind myself. I remember that process and I don't remember -- I would be very grateful to see that again, as we revisit this.

CO-CHAIR PINCUS: Raquel, and then David, and Mike.

MEMBER MAZON JEFFERS: So I think my questions are super concrete at the moment. Exactly -- I just think I need some clarification exactly for this proposed measure, how is it administered and what is in the numerator and denominator. So is it the entire CAHPS survey that is administered to a random sample of members in a health plan and then the measure is calculated by pulling out the responses to these items in the ECHO? So that is one question because if I am -- because there are two other clarifications that I am still struggling with. One is at one point you said that you would only be administering the ECHO to people who have received behavioral health services and it is my understanding that one of the things you are trying to get at is how ready the access was to

the behavioral service so it would make sense then to administer this survey to people who -- to everybody to see if they needed the behavioral health service and then, therefore, didn't get the behavioral health service.

And then the last thing is that I have heard two different things, one that the measure can be used in primary care settings but if you actually read the PA that was prepared for today, it says that the denominator exclusions include people who receive behavioral health services only in primary care settings.

I'm sorry, could you just nuts and bolts help me understand those three things?

DR. CLEARY: Sure. The instrument was designed, the existing ECHO was designed for people who had received behavioral health services identified in some way by a health plan or Behavioral Healthcare Organization. So, only people who had received. That was what the design was.

What I talked about more recently

testing and developing was we perceive a much broader need for questions about access and availability in a general population. So we actually are testing questions like that now in Connecticut. Our belief now based on the focus groups we have done and work with different constituents is that it would be desirable to have both a survey for people who have received behavioral health services and a set of questions that are applicable to a general patient population addressing the questions you have asked.

And when we originally developed it, it was most people wanted to use it only for people who had, in addition to primary care, received specialty services. So that was the definition of the denominator. But it does ask questions that relate -- it could be used in people through a primary care panel because as we all know, at least many of know, a lot of behavioral health services, there is a high proportion of primary care patients with behavioral health issues. A

large number of those patients are not recognized. And those who are recognized don't get adequate treatment. So that is an important subset of people.

CO-CHAIR PINCUS: Dave.

MEMBER EINZIG: Hi. So my question has to do with concept of lumping. I am a child psychiatrist and pediatrician doing a collaborative care model in an urban setting and see lots of patients. So I am having trouble with lumping all people with mental health together in the same group, not just with substance use disorder and mental health but within mental health itself. So, for example, preparing a person with high stressed family, and ten kids, and fetal alcohol, and poor executive functioning, very poor, a lot of stress, comparing that to a person with anxiety and an IQ of 140 doesn't really makes sense to me.

And then even within the high needs group, where they may need more intensive supports and services and frequent visits, some

of those families may have supports embedded, case management services, and home services, PTA providers and whatnot, and comparing that with folks who don't have those supports.

So, just help me understand how one evaluation to folks with mental health services, how that can be more useful.

DR. CLEARY: Sure. It's a great question. I would just reiterate the role of the CAHPS Consortium is to develop a way of assessing care experiences. And we don't dictate that it be used in a particular way. You know we tried to define a numerator and a denominator for precision but what typically happens is let's say I, as an individual rather than a CAHPS representative, starts talking to you about how to evaluate your services. You make the statements you just made. I would say it sounds perfectly reasonable to me. What would be in your opinion the appropriate groups to compare let's say across providers? So let's say you say well, I am really worried about severe and

persistent mental health problems. I say okay, can that be defined. Can we compare those across what subsets of providers and such?

What I am saying is that is basically an evaluation design issue that I don't dictate. If the State of Connecticut uses this, what they have told me is they are interested in general, broad access. Maybe on the next iteration they would be particularly interested in population X or Y. Because inevitably when you do the first assessment, you say gee, I don't have enough precision to answer the questions that you just asked. Is there a better design for providing information to your system or your providers about how well you are doing with different subsets of patients? That is really a design issue that you know in part of my life I do a lot of work but as a CAHPS developer, it is not what we dictate.

CO-CHAIR PINCUS: Kraig, is your card up?

Mike.

MEMBER TRANGLE: I have I think one question and two or three comments that will be kind of quick.

One is that in my part of the world, which is very similar to David's, integrated care is becoming more and more the mode, instead of the exception to the rule. And to sort of say it would be a rare patient that, in our system, it would always start in primary care. Somebody might try and treat them for depression. They don't do well, they go to behavioral health. And they may see the behavioral health provider in primary care or they may see the same person in a separate behavioral health clinic.

When you are excluding people seen in primary are, it strikes me as that it is going to become increasingly problematic. And then you will get this thing about only seen in behavioral care, or only seen in primary care, or both. As the world is evolving, I just think you should rethink that in a way so that it is both practical and useful of how you are subdividing.

DR. CLEARY: That is a good comment and I agree with you. And as Harold and some of you know, I started my career looking at the recognition and management of mental health problems in primary care. So, I totally understand what you are saying.

MEMBER TRANGLE: You know and the same thing pertains for substance abuse.

DR. CLEARY: Yes.

MEMBER TRANGLE: As we are dealing with the opioid epidemic, we are having both behavioral health and primary care doing Suboxone and having people do kind of supportive therapies in primary care. It is just not distinct.

DR. CLEARY: I agree. And to be honest, that was part of the original specification that we maintained and I think that was dictated by some groups we were working with but I agree with your observations.

MEMBER TRANGLE: And then you said something that maybe I misheard. But I think when you were talking about CG CAHPS and HCAHPS

you sort of said it is not meant to somehow give feedback on the individual. Did I hear that correctly? Individual clinician, you know.

DR. CLEARY: Oh, no, I meant individual patient.

MEMBER TRANGLE: Individual patient but individual clinician --

DR. CLEARY: Now it is true, also, for -- there is also a huge issue with individual providers in that relatively little care is provided by an individual provider but Clinician Group CAHPS is designed if one wanted to assess clinician care.

MEMBER TRANGLE: You know my comment is, at least in my part of the world, I don't know who one is. I mean sometimes it is a health plan. Sometimes it is an employer doing contracts where patient satisfaction is important. But it is probably getting more legs and more use for that than the group level.

And you know even most sort of say you have to have an n of 100 or more to think that it

is at least somewhat statistically plausible but it has been used that way a lot.

DR. CLEARY: I understand but, for example, HCAHPS has been used like that and --

MEMBER TRANGLE: And our system is --

DR. CLEARY: -- and CMS is really trying to push back. It is not appropriate to use an HCAHPS score to make an attribution to let's say the admitting physician because it asks about the hospital experience. You have to very careful about the design to make attributions to individual physicians, although I agree with you many people are eager to do that.

MEMBER TRANGLE: And the focus is on the questions asking about did your doctor listen, did they explain things intelligently, understandably.

DR. CLEARY: Yes.

MEMBER TRANGLE: And then my last question is kind of a half question, half comment but we have been struggling to help ease people sitting in our hospital psych beds that are

always full with people flooding the ERs. We have developed transitional housing to get them out, crisis beds, and now have our second or, depending on how you define it, third group home coming online. And when I try to get data about can we look at patient satisfaction in that realm, and the same thing plays out for quality actually, too, it feels like I don't even know where to start. I do literature searches and you can see the Army has some things that they have out there on their website and this and that but it feels like it is the Wild West. And I don't even know where to start of how to -- you know what are good surveys to use for patient satisfaction in those zones. Do you know anything about that, group homes, long-term residential treatment centers, that kind of place?

DR. CLEARY: I mean I know about it. I don't have experience with our instrument. We are doing a lot of that work with CMS. I mean it is not the same but analogous work of like

assisted living, different types of facilities. So there is a whole family of instruments. We haven't done that in behavioral health services but it is a good point.

CO-CHAIR PINCUS: Okay, so Tracy said we have somebody from the public.

DR. LUSTIG: Hi, so we did receive a public comment that was sent to us yesterday that couldn't stay. So I just wanted to read it and Paul, also give you some feedback.

It is from D.E.B. Potter who says that she is speaking as an individual, not as a representative of HHS or ASPE and she notes that since 2011 she has been an ex officio member of the MAP Dual Eligible Beneficiary Workgroup.

And she writes the CAHPS behavioral health experience with care measure ECHO is like several other CAHPS measures included in the MAP Dual Eligibles Beneficiary family of measures.

In 2015, the MAP Duals Workgroup undertook a measure alignment exercise to identify the users of measures in the Duals

family of measures. Based on the 6/25/15 version of the MAP tool, which is no longer on the MAP site, several of the States involved in the Medicare-Medicaid Financial Alignment demonstration use ECHO to assess performance. NQF obtained data on ECHO's use by abstracting information from the Memorandums of Understanding signed between CMS and the states. In 2015, the following states were using ECHO and these are all capitated state demos in California, Illinois, Massachusetts, Michigan, New York, Ohio, South Carolina, Texas, and Virginia.

And she says as of 2017, all of these demos are ongoing, although Virginia will end this year. She also noted a Commonwealth Issue Brief from march 2014 which further identified ECHO and the CAHPS plan measure as CMS core measures for the demonstration.

She says that in the worksheet that was done for this measure it is noted that no recent data on performance results were provided for the 17 PRO PMs included. And she says based on

ECHO's required use in CMS-funded programs that involve multiple health plans, I suggest that perhaps some recent data from multiple years do exist for the ECHO items, just not easily seen or obtainable by the public and/or the research community.

She says given the importance of having NQF-endorsed PRO measures for the behavioral health population, I, as an individual not representative of HHS or ASPE, urge the Behavioral Health Standing Committee members, the ECHO measure steward, and/or the NQF behavioral health staff to reach out to CMS MMCO, these states and their health plans to determine if more recent PRO item data does exist and to request that the data be submitted to the committee for evaluation. And she also says that we could begin the conversation with Alice Lind, who is a member and former co-chair of the MAP Dual Eligibles Workgroup.

DR. CLEARY: Wow. I didn't know any of that --

DR. LUSTIG: And Paul, we can forward you the information and help get those started.

DR. CLEARY: Yes, if you could forward that to me. Obviously, or I don't think, that is the role of NQF but we would delighted to try and -- we work very closely with CMS to try and get those data and do the kinds of analysis that would be informative.

DR. LUSTIG: Yes, we are happy to help you make those connections. And we literally just got this last night ourselves and so we would be happy to share that with you.

DR. CLEARY: That would be great.

CO-CHAIR PINCUS: And I would also suggest, Paul, that you also contact Kate Goodrich and Jeff Buck and her group who were involved with the Inpatient Psychiatric Facility Reporting Program and looking different opportunities for PRO kind of and perceptions opportunities.

Peter, and I think you are the last one.

CO-CHAIR BRISS: Do this has been a

great conversation, thanks to everybody.

Paul, I have one more piece of advice for getting through the NQF system. So as I listen to you talk, there is sort of a current state of the measure and there is sort of you are frequently leaning forward into potential improvements, or tailored applications, or more specialized things. And for the purpose of getting through the NQF process simpler is better and what we need is the off-the-shelf generic version that we can identify today -- that we can evaluate today.

And the forward leaning work is really good but it makes hard on the committee to kind of get a sense of where the current reality is. So this is my other piece of advice that I didn't give you before was I would focus just on the generic version for this audience.

DR. CLEARY: No, I totally agree. I was just giving you my thinking because I don't get to talk with a group this knowledgeable very often. So, I thought I would share sort of our

thinking and the input has been incredibly helpful. But I understand that discussion and the approval process are very different.

CO-CHAIR PINCUS: So Paul, thank you so much for your time and for your expertise. Please, we would like to continue this conversation and get feedback on an ongoing basis about progress with this measure. And please, with your colleagues at AHRQ, and CMS, and the CAHPS group, convey our enthusiasm for sort of moving ahead with this.

DR. CLEARY: That's great and thank you very much for everyone's time. And if you could forward that comment, that would be great.

And I will just make an ask. As we move forward, we are very interested in input from technical advisory panels, experts, et cetera, et cetera, and we have done that in the past. If anyone is interested informally, with very low burden, providing comments or input on the development issues, not the application but the development issues as we move forward, I would be

delighted to get your information and keep people in the loop because we have gotten some incredibly useful comments this morning that I really appreciate.

CO-CHAIR PINCUS: Okay, terrific. So, for the committee, we are going to take a break and get back together at 10:15.

(Whereupon, the above-entitled matter went off the record at 10:03 a.m. and resumed at 10:16 a.m.)

CO-CHAIR BRISS: So, everybody, could we reconvene, please?

So the next measure is Continuity of Pharmacotherapy for Opioid Use Disorder. I will keep trying this until it works. The next measure is Continuity of Pharmacotherapy for Opioid Use Disorder. And it is customary to stop other conversations and rejoin the conversation. And the measure steward is RAND. Would you like to introduce yourself and take three minutes or so and key up the measure for us, please?

DR. MATTKE: Okay, thank you, Peter.

Hi, Soeren Mattke at RAND. I am a senior scientist in the Boston office and I am the project director for this development exercise. I am here with Liz Sloss, who worked with me closely. She is an epidemiologist.

And I always say I am a cardiologist by training. So anything north of the carotid artery is kind of a mystery to me. So we have Kate Watkins on the phone here from Los Angeles. She is actually a psychiatrist and knows what she is talking about. She was working on this project as well and will help me when the questions get too clinically technical.

Okay, so great to be here. We are discussing this opioid treatment continuity measure first. As you all know, opioid abuse has become a national public health crisis. It is now killing on average 90 Americans each day, about four times as many as 15 years ago, which makes it on par with quite common diseases like chronic liver disease or motor vehicle accidents.

In a 2014 survey, more than ten million

people reported prescription opioid misuse, which puts them at risk for both overdoses and then subsequent heroin addiction. This dramatic increase has brought the treatment of OUD to the top of the agenda for clinicians and policymakers. The Surgeon General, for example, recently mailed a call to action to over two million clinicians and the governors of many states, red and blue, have prioritized improving access to prevention and treatment.

Pharmacotherapy for OUD is an effective but underused treatment option but we expect that to change for three reasons. First, we have now several FDA-approved drugs for use in regular clinics and offices, which facilitates treatment greatly compared to methadone, which, as you know, can only be dispensed in specially licensed treatment programs.

Second, leading professional societies and SAMHSA have issued support of statements to promote OUD pharmacotherapy.

And third, national quality measurement

and reporting schemes such as HEDIS and the Joint Commission set are capturing or will capture initiation of mitigation-assisted treatment.

But as the evidence in our application shows, there is actually a deadly risk with prescribing opioid pharmacotherapy treatment without ensuring continuity of care. Pharmacotherapy for OUD is not like putting patients on statins. Both the agonists, the partial agonists and the antagonists alter a patient's tolerance of opioids and thus, there is a risk of overdose when treatment is discontinued.

As we have shown, that risk was up after about three days of treatment interruption and persists for at least two weeks and that finding is quite consistent for the different opioid medications.

For that reason, we have proposed here a treatment continuity measure that capture where the patients on pharmacotherapy for opioid use disorders remain on treatment for at least 180

days and have no gaps of greater than seven days. We hope that this measure will encourage providers and health plans to institute adherence interventions when starting pharmacotherapy in order to keep this effective treatment option safe.

Our data show that improvements in continuity of care are both needed and possible. The average measure score at the health plan level was around 25 percent and ranged from close to zero to over 60 percent, suggesting both room for improvement and best practices to emulate. We provide expert panel support for validity and usability of the measure. We provided data on reliability testing and expect this claims-based measure to be feasible to implement.

I am looking forward to our discussion. Thank you.

CO-CHAIR BRISS: Thank you. So, we will move to importance of measure and report. Would one of primary reviewers like to tee this up for us, please?

MEMBER GROSS: Is the importance of the measure is that --

CO-CHAIR BRISS: Importance of measure and report. So, we will -- we are back to the drill that we were doing yesterday. So, we will run criterion at a time. So, the first criterion is importance of measure and report.

MEMBER GROSS: Just a quick question to -- Charlie Gross with Anthem.

The database you based it on included commercial Medicaid numbers as well?

DR. MATTKE: No, it is only commercially-insured employees of large and self-insured employers.

MEMBER GROSS: I still think it is a very important measure but I think maybe -- and this may be more of an NQF sort of discussion that we can save for later. I think increasingly, as government business is a bigger part of the payer mix, particularly with disorders such as this, that the need for those populations to be part of the measurement data

set is critically important.

Having said that, I think this is an incredibly important measure.

DR. MATTKE: And we do intend to test it on Medicaid. Actually, we have the data. We also intend to test it on Medicare patients for which we couldn't get the data in time. It is just this call came out with such short notice that the best we could do was commercial claims, which is the easiest to secure.

MEMBER GROSS: Sure. Thank you.

CO-CHAIR BRISS: So, Raquel?

MEMBER MAZON JEFFERS: I also am very enthusiastic about the importance of this measure. So thank you for bringing it forward to the committee. I just had a couple of questions.

One is that the measure is focused on the use of medication-assisted treatment to support opioid use disorders alone and doesn't include any counseling component as part of the measure. There is a lot of evidence to support the increased efficacy of medication-assisted

treatment coupled with counseling.

I know this is a matter of debate in the field. I was wondering if you could talk a little bit to why you chose to move forward with medication alone as part of the measure and leave out the counseling component.

I am going to hold off on --

DR. MATTKE: Okay, this is really a practical issue. With medications, you can basically track whether a patient adheres to the treatment plan because the claims tell us how many days were supplied and how many refills a patient did or how many injections a patient got. There is no normative source to tell us how many counseling sessions you ought to have to accompany your pharmacotherapy. Like some patients may have to see a counselor or psychiatrist once a week, even twice a week. Some may go in once a month. Others might receive the counseling component through sort of non-medical settings like self-help group or churches and we wouldn't see those in claims.

So, we had long debates in the team whether we could also measure continuity of counseling but we couldn't find a way of operationalizing it for measurement purposes.

MEMBER MAZON JEFFERS: I would just -- I can understand the practical constraints that you are struggling with but there are criteria based on ASAM patient placement criteria based on the level of severity of the patient for a recommended level of counseling intervention that are scientific and consistent. So based on someone's level of severity, there are recommendations for the level of intensity of their counseling that are standardized.

DR. WATKINS: I think -- this is Kate Watkins. I think the problem is that getting that out of claims data is difficult because a person may see -- first of all, it is really not clear what counseling means and what is the -- how you define it. And getting that information out of claims data, whether or not it is a physician who is just counseling the patient in

the process of when they are prescribing the medication, that could count but that wouldn't be necessarily captured in the claims data.

CO-CHAIR BRISS: So, Mike.

MEMBER LARDIERI: Yes, I guess I just need to understand why you wouldn't be able to -- I can understand if they are going to a church group or something like that but every other outpatient service, the outpatient providers would be providing a claim to get paid. A psychiatrist would provide a claim to get paid.

If they are not doing just medication, the E and M codes, I believe show that you are doing counseling or not. So I am having a hard time understanding why you wouldn't be able to get it under the claim. Because it is so important to have the counseling component as well as the medication.

I'm just looking through the ASAM. Do you know what the ASAM criteria states around that, whether you have to have counseling as well? Because so many of behavioral health state

programs are requiring you to follow the ASAM criteria. So have you looked at that, what that says?

DR. WATKINS: We have. It doesn't define what counseling -- it doesn't say is it -- and it is actually psychosocial treatment. So but, again, it doesn't define whether or not it is contingency management, whether it is relapse prevention, whether it is just your physician talking with you.

And yes you could use the E and M code to code for delivering the medication and talking to the patient but, again, you really don't know what is happening in that interaction.

MEMBER LARDIERI: And maybe it is not that important what is happening, the importance is that it is happening I think is the way I would look at it.

DR. MATTKE: Again, if you had more content on the interaction, you could do that but if somebody has regularly, say primary care visits where increasingly addiction treatment

takes place, we really have no idea does the patient talk about the addiction or is this about diabetes.

I understand sort of conceptually what you are trying to do, it is just from a measurement perspective, it is extremely difficult.

MEMBER LARDIERI: I'm still having a hard time with that because we are so much integrated now and we need to bill for both sides of the house. And even many Medicaid programs, now this is just on the commercial side, but on the commercial side you pay for both a physical health and a behavioral health payment on the same day. So I am still having a hard time why you wouldn't be able to capture that claim. I don't get it.

CO-CHAIR BRISS: So I think I would like to suggest that several of us had a variant of this kind of comment about the medications work, the counseling works, they are likely to work better together. This isn't a part of the

measure we have in front of us today.

And so for the future, you guys might want to think about whether at least there is some minimum standard. You can never get a perfect measure of the content of a counseling session but you might be able to do some minimal standard of whether some interaction happened that might be better than nothing. And that is likely to be a future-looking thing about the measure that is in front of us.

So, Shane.

MEMBER COLEMAN: I am just curious if you guys could comment a little bit on the six month time period that it shows. I know you mentioned preventing overdoses when people come off the medication, though I think that occurs at any time, like you mentioned. I think you guys quoted three days.

And then also, trying to improve people's outcomes or, I am guessing, maintain sobriety. And I know that you mentioned methadone maintenance having maybe higher rates,

I'm not sure of maintaining sobriety but, of retention of certain time frames. But I am just curious for other MAT medications, non-methadone medications, I'm curious why the six months.

DR. MATTKE: We don't say just six months. We say minimum six months. So, if a patient stays on MAT longer, we would still say give me the risk of a lethal overdose if you interrupt. You still want to look for the gaps in treatment. We do not advocate putting somebody for just six months but I think that is kind of a lower bar in terms of a sort of reasonable duration of treatment that you can say somebody is actually in OUD pharmacotherapy as opposed to somebody just gets initiated and then there is no follow through and the patient drops off the radar.

DR. WATKINS: There is no question that longer is better. I think the evidence is really clear across the different medications that when you stop, you do worse.

But given that, there is also empirical

data about how long is sufficient. And that is why we ended up choosing six months because of the FDA. The FDA approval trials were between three and six months. Because there is no empirical evidence about whether it should be six months, nine months, twelve months, two years, five years, forever. I think that is an area that is in need of more research.

And we elected to go with the six months because it should be at least six months.

MEMBER COLEMAN: Okay, thanks. Yes, I guess I just bring it up because it is almost in any venue you start thinking about offering MAT. That is a question that comes up. So I am just kind of curious at your guys' rationale. So, I appreciate it. Thanks.

MEMBER ZUN: So I have a couple of questions. So the first one is isn't the ultimate goal to get them off of any medication? So that is my first question. And how would we then -- or maybe they transition to some other therapy. And I am seeing no. Okay.

And then the other thing is if I initiate treatment in the ED, who would that apply to this? Would there then be that -- if they did -- if they continued for six months, then it would apply but would it be -- would it come under my -- would my hospital be dinged if they didn't follow through? Do you follow that one?

DR. MATTKE: So the second one is easier to answer. We are currently calculating on the health plan level partly because the attribution is actually tricky with these patients maybe being initiated in like a detox center and then going back into primary care. So we haven't figured out attribution to providers. We just say okay, a health plan who has the sort of totality of the 30,000-foot view of the patient's care should work with the contracted providers to ensure treatment adherence and continuity.

The first one, whether the objective is ultimately getting patients off any substance I think, Kate, I defer to you on what the evidence

says.

DR. WATKINS: I do not think that is the goal, necessarily of treatment. I mean obviously, it is up to the patient and you have to think about what the patient's goals are. But in someone with diabetes -- I mean typically, opioid addiction is thought of as a chronic illness. And you would never say to a person with diabetes the goal is to get you off medication. People do better when they are -- the longer term treatment -- outcomes are better with longer term treatment.

It may be that you have a patient that wants to be off medication and that would be fine and you would work with them on that but for many patients, they stay on methadone for years and years and years and that is what allows them to have productive lives.

MEMBER ZUN: I'm not going to dispute that but I am going to dispute the diabetes issue because the goal is if their diabetes is because they are overweight or they have an inappropriate

diet, you try to control that. So why would this not be -- you know we want the least difficult therapy. And if going to Narcotics Anonymous is therapeutic for them, why would we want them to continue taking methadone the rest of their life?

I am just not following all the reasoning.

DR. MATTKE: And we are not saying they should. We are just saying as long as you are on methadone or other substitution treatment, you should not have treatment gaps because that puts you at risk.

CO-CHAIR BRISS: And it looked to me like, just watching the body language in the room, there appeared to be a lot of body language suggesting that people feel like longer term treatment, at least for some people, is a reasonable thing. So I don't know that we need to dispute the aptness of a diabetes analogy.

Next.

MEMBER PARISH: I just sort of had a comment that was maybe something similar to his,

along the line that when is this going into effect because I think that there is a large culture of doing detox and the moving them off into rehab without any medications.

In my work as a medical director I cannot tell you how much I get well they need two more days of detox because they still are on two milligrams of Suboxone. And I am like, that's fine, move him to rehab. But I think there is really a culture there.

And the other concern I have is especially for the rural areas, in terms of the Suboxone prescribing, we do not have the prescribers. And as you know, the DEA requires at least the first visit to be done in person and is there anything that can be done in terms of looking at having that changed so that more can be done through telehealth, et cetera for the rural communities, where it would get hit the hardest, I think, with this quality measure?

DR. MATTKE: Yes, very important questions, well beyond what we can do with our

measure. I think we can just point out these gaps and discontinuities. What to do with it are probably people, policymakers, and other decision-makers.

MEMBER TRANGLE: I'm a Suboxone prescriber. I don't do it a ton. And I do think there was something to what was said in terms of when I have somebody coming to see me, one of the first discussion we have is where are you coming from, as a patient. And some people will come to me and say well I want to go through detox and go to NA and do some of that kind of stuff. And the focus ends up being sort of an induction to get them off. And if they really want that a lot, I am willing to give them a try always.

I know that in reality most people that are truly dependent are going to do far better on a methadone program or a Suboxone maintenance kind of program. But I think it is totally within the scope of good practice to try an induction and see if they can get psychosocial supports and do okay.

And it doesn't seem like your measure allows for short-term use, intending it to be short-term use and then stop it. So that if I do an induction and they just want to get off it, I will get dinged for not keeping them on it for six months when that was never their plan, unless I am misunderstanding the inclusion criteria. So that is one question.

I have a couple other questions but do you want to just answer that one first?

DR. MATTKE: So there are initiation measures like NCQA is in the process of adding MAT to its initiation and engagement measures, which would capture your patients.

MEMBER TRANGLE: But would they automatically in or are they excluded, someone where it has never been the intent to do maintenance?

DR. MATTKE: We would require two prescriptions before we can find somebody. So, depending on how long your prescriptions are, we may or may not capture that patient.

MEMBER TRANGLE: Okay.

DR. MATTKE: The question is can you tell after a very short period whether or not a patient is sort of well controlled and stable under a Suboxone regimen or does this indeed need the 180 days.

MEMBER TRANGLE: Sometimes yes, sometimes no.

A couple of other questions. Well one is really a comment. When I went through this, I was on the committee or the three or four, or now it is five or six people that reviewed these. And one thing that struck me was that the extended release naltrexone, the Vivitrol, only had one study, 250 people in Russia. And it wasn't graded.

But as it turns out, my qualms about that sort of subsided because there was a New England Journal article that came out March of '16, the third week, where it was -- I jotted some notes. Vanita, this is why we get along so well -- where basically the median time to

relapse was twice as long for people that were on Vivitrol, 10.5 versus 5 weeks. And there were about a little bit less than 50 percent more relapse events during the six months of the active treatment. It was followed for 78 -- six-month intervals but followed for 78 weeks. But six months was active treatment with the Vivitrol.

So there was at least another article sort of saying yes, Vivitrol probably does work. Otherwise, I was going to hassle you about that but I won't.

DR. MATTKE: It is a little bit a quirk of the NQF process in the way that they ask you to present evidence. The first step is like are there any nationally accepted guidelines that support what you are trying to do here. And we had the VA/DoD guideline which basically has the -- which is the approval study for Vivitrol.

So then the question is so show the evidence that is behind those guidelines and explain what it does, which means then we go to

the Russia study which was the approval study. And yes there are, not just the New England Journal article but there is other papers in real world settings that looked at efficacy of Vivitrol.

MEMBER TRANGLE: So my last question is I think the measure goes to 65. And if I am 66, I am out, right? And I am just you know is that because you didn't have evidence but it could have been a good idea to do it later or just tell me what your thinking was behind that.

DR. MATTKE: No, that was just a time constraint. In order to test it in patients over 65, we would have had access to Medicare data plus Medicare prescription plan claims.

MEMBER TRANGLE: Oh, you didn't have the Medicare data so you didn't include it.

DR. MATTKE: So we had like three months to prepare the submission and especially Part D data, actually, not easy to get. And so it was just a practical perspective and we completely intend to test both Medicaid and Medicare going

forward.

CO-CHAIR BRISS: So you have heard a couple of times now that broadening the patient base in a variety of ways might be constructive going forward.

So, Rhonda, I would love to get through the remaining comments on this criterion relatively quickly and be starting to think moving toward a vote.

MEMBER ROBINSON BEALE: I have a patient safety question. Has there been any studies to look at patients who have been on medication-assisted treatment, dropped off and looked at that impact on the death rate?

So in other words, is this increasing the number of people who are dying of overdose or is it the access to the opioids in the streets?

DR. MATTKE: No we actually show quite a bit of evidence the mortality goes up if you interrupt medication-assisted treatment. And that is because if you go on agonists, partial agonist or antagonist therapy, your learned

tolerance of synthetic opioids, particularly, decreases. And then if you relapse and use kind of what you used before, you have a higher risk of overdosing.

So the safety issue is not the treatment. The safety issue is interruption.

MEMBER ROBINSON BEALE: Understood that the treatment is making patients more sensitive to that. So the reason why I raise that -- and, believe me, I am for medication-assisted treatment, I always get concerned when we put a measurement in place, and particularly if it is going to be used in any way to judge the efficacy of treatment if there is a huge safety factor that is not written into this. There is nothing that warns people on this and also that there is built into, and I don't know, maybe you can't talk, but in the guidelines something that helps providers to understand and that they need to be very vigilant about counseling people on that, as well as their families. I just want to make sure that people are aware of that.

So, I see that as a huge safety factor.

DR. MATTKE: Yes, and that is precisely why we want that measure because the NCQA measure will look at initiation, how many people get put on MAT. The Joint Commission measure already does that but no measure, thus far, looks at continuity, which is exactly what you point out that unless you are also looking into continuous care, you are creating a safety issue.

MEMBER ROBINSON BEALE: Do we know that if the treatment stops at 180 days that the incidence of overdoes goes down? Is there a time frame to this?

DR. MATTKE: They are not saying you should stop at 180 days.

MEMBER ROBINSON BEALE: I understand.

DR. MATTKE: So if you say this is the minimum duration for which we have evidence, we fully acknowledge that the duration could be years but we say for the duration of your pharmacotherapy, you shouldn't have these gaps because these gaps are dangerous.

MEMBER ROBINSON BEALE: Understood. I'm just to respond to the lay provider public who will try and do this, even if they become Suboxone prescribers and others, need to make sure that the safety factor is being addressed in some way because the measure is going to promote the use, it is going to promote the use of 180 days. And unless the measures only apply to very specific types of prescribers and not more generalized, that is where I get more concerned.

I don't know if I am making my point.

MEMBER PINDOLIA: Hi. So this question is we had a huge discussion at our PNT like in March last year because of all the national considerations on re-reviewing all the different possible drugs that could be used, what is the pros and cons, what is going on. And even after removing all stops and making this publicly acknowledged that please, if you have a patient, these are available, with the only exception as being Evzio because it is $4,500 a dose of having a little bit more criteria for that.

My question for you is with all this, we are just completing a one-year post analysis to take back to PNT. And it is like to your point, after the first dose, there is very few that come back for their refills. So, in your evidence for wanting to have the six-month continuation, did you run across the data of what makes that discontinuation?

And the reason I am asking is it goes back to the earlier comments about the counseling. I think there was an NCQA depression measure a while ago. And within the first six weeks you had to have a follow-up phone call and it was sort of linked with also drug continuation. And it would really help for a health plan who is trying to remove all the blocks and trying to have the access of the drugs available, trying to promote this in the community but if they are not having a physician or a someone to follow-up, even if it is a PCP, there is nothing that we can do as a health plan. And maybe if it was part of a quality initiative

to say the measure requires both, that is a different message, as a health plan, we can send.

So, two questions. One is when you are reviewing for this, did you come across the evidence of why people are discontinuing? And then second, to reconsider of having it coupled.

DR. MATTKE: Yes, Kate --

CO-CHAIR BRISS: On the second one, I think we have essentially talked about that already. So, I would like to suggest you answer the first one and then we go on, please.

DR. WATKINS: We don't know why people are discontinuing. We think that at least some of it has to do with transitions in care, where they are transitioning from one care location to another and they are fall between the cracks. But we don't know why people drop out.

CO-CHAIR BRISS: Thank you. Raquel?

MEMBER MAZON JEFFERS: I wanted to go back to Rhonda's questions about safety. So this is not a scientific study but based on experiments -- based on experience, in New Jersey

we are seeing the highest rate of overdose immediately post-NARCAN reversal for people. So, individuals who have overdosed, they have entered the emergency room or they have had an encounter with a police officer who administers the NARCAN has been reversed but they have received no engagement in treatment and they have not received any medication-assisted treatment, which if it is a medication like Suboxone, it includes a protection against overdose.

So the death rates we are seeing are more to do with the chronic relapsing nature of the disease. After the NARCAN administration, people feel very uncomfortable. They go back on the street and they are very driven to use.

So I see the continuity of care using a medication like Suboxone for people with an opiate use disorder and maintaining them for six months, hopefully linking them with counseling to begin to really address their underlying addictive disorder is really increasing safety because you give them a chance to actually

experience what they need and to grow and develop and be able to achieve some recovery that would prevent them from using right away. You are also administering a medication that has an overdose protection quality.

So, we are actually seeing more overdoses for people who are, as Brooke described, going into residential treatment with no medication, being released, and then using upon release because they are still feeling withdrawal and craving. Their brain functionality hasn't been normalized and then they relapse and use. And that is where we are seeing the overdoses happening.

CO-CHAIR BRISS: Thank you.

Tami and then I will give Shane the last word and then I would like to move to a vote, please.

MEMBER MARK: Okay, I just wanted to add a couple of data points to the discussion. I think the main paper people cite to show that adherence is important is this Weiss paper that

followed people in my study that showed that when people discontinued prematurely, there was a huge relapse rate.

They published a study recently where they continued to follow people for 42 months. And after 42 months and about a third of the population was abstinent and not on an agonist therapy, meaning that after a significant amount of time, three years or more, you can get a third of the population into recovery and they don't need buprenorphine. So it is not like diabetes.

But that is not saying that if they are not on it for the first nine months, there is not a high risk of what Raquel is talking about. So, I just wanted to bring that data point up.

The other data point is on the safety point, I believe the guidelines don't say that buprenorphine is contraindicated in pregnant women. I didn't see that as an exclusion.

And then the third point is when I have done these -- I know you excluded methadone because you said it is not available on drug

claims but when I have used claims, I can identify methadone through a HCPC code. So I think you can include methadone in your specifications and that would be important.

DR. MATTKE: Yes and we do include methadone if it is given through the HCPC code H0020. We do not include it if it is dispensed through a pharmacy claim because, legally, that is not permissible.

MEMBER MARK: Yes. And then just my final picky point, and this goes to the evidence, I mean your evidence shows that the adherence rate is low. You used a claim period, though, right? Because I didn't see that you used a claim period. These are all people who were newly initiated on the medication. You are not picking them up at --

DR. MATTKE: It is cross-sectional. So, we did not specifically look for new initiation. We basically start looking when we see the patient.

MEMBER MARK: Yes, see that is

problematic. That is not what this measure is. This measures is newly initiated and followed for 180 days, right?

DR. MATTKE: No, it is not.

MEMBER MARK: Well, some of these people could have been on it for five years and --

DR. MATTKE: Correct.

MEMBER MARK: So, that is not the right measure.

DR. MATTKE: I think the really critical component of the measure is the gap, not necessarily the duration.

MEMBER MARK: Well your measure is not a gap. Your measure is a duration. Your measure is whether people are on it for --

DR. MATTKE: In order to pass the measure, you have to have at least 180 days in our observation period and no gap greater than seven days.

MEMBER MARK: So you are including people who could have been on this for three years and then they just stopped and so they

wouldn't have 180 days because they just were in the last month of the three-year period on the medication.

DR. MATTKE: We could, yes.

MEMBER MARK: That is very problematic.

I mean that is not the way the measure is described and I don't think that is an appropriate measure. The measure is -- we want to get people who are newly initiated because that is the critical time period. That is what the evidence shows, if you are newly initiated, you should be on it for nine months, unless there is a safety issue. But that is not what this measure is.

CO-CHAIR BRISS: So let me suggest on this point we have had a pretty far reaching discussion on this point. This is really a measure spec thing that I would call reliability.

So I would sort of like to vote on --

MEMBER MARK: Well I think it -- let me just say, Peter, it does go to evidence because the evidence that they are presenting of a low

adherence rate is based on this spec that is inappropriate. So I would argue the adherence rate is probably much higher if they had the appropriate spec. So I might say that when we vote on evidence, we might want to say it is insufficient.

DR. MATTKE: So to be clear, we are aware that many patients are on long-term treatment and should be. And what we want to measure is primarily the gap in care because we know that that is predictive of mortality. There are other measures of treatment initiation, which we didn't want to duplicate.

DR. WATKINS: I would also say that it is unusual for someone who has been on long-term medication-assisted treatment, it would be unusual for after a number of -- or it would be a minority of people after multiple years to decide to stop.

MEMBER MARK: That is just the opposite of the evidence I just cited from the study. It showed that a third of the people after three

years stopped and were no longer addicted. So that is what the evidence from the Weiss study shows.

CO-CHAIR BRISS: Aren't you saying the same thing? So a third of people is a minority of people, right?

DR. WATKINS: That's correct.

CO-CHAIR BRISS: So, Shane, I will give you the last word and then we will vote on evidence, please.

MEMBER COLEMAN: I think I just wanted to I guess comment on the retention rate issue again, like understanding a little bit about why people might develop more than a seven-day gap. I mean I guess I just want to say how important it is, again, because it is mysterious. It is something we are struggling with too in our system how to keep -- you know if you look at the studies of retention rates might get as high as 60 or 70 percent. Some of them are seeing people like every other day during the first week. Like in our induction, we have them come back within a

week. So, we are playing with factors like that but our retention rates -- I won't go into too many details. Like we are also, because of our priority systems in the state, focusing on IV users and things like that so we don't get the lower end of the spectrum as much but our retention rates are like 20 or 30 percent. So we are really struggling to figure that out, too.

So in this measure, I would be -- I mean it is still good. I think it will prompt people to continue to think about this but I am little uncomfortable that I feel like there is no answers out there for how to improve your rate. But, again, maybe it will drive us to think about that. I don't know.

And real quick, the natural course for folks, in my experience, is that around two or three years, they do start asking should I stay on the medication. It is not unlike other chronic diseases. And I would argue that at least the majority of folks actually ask that question. Not all of them will feel comfortable

getting off but I still actually think I feel very comfortable with in my experience the majority of folks, after if they make it three-ish or so years to really start thinking about coming off the medication.

CO-CHAIR BRISS: All right, so I would like to move us to a vote. This is on the -- it is essentially on the evidence of the measure concept, which is the continuity of treatment without gaps would make people better off.

MS. QUINNONEZ: We are not voting on Measure 3175, Continuity of Pharmacotherapy for Opioid Use Disorder. We are voting on evidence.

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

Okay, voting is now closed. For Measure 3175, the results read as follows: 17 percent voted high, which is three individual votes; 56 percent voted moderate, which is ten individual

votes; zero percent voted low; and 28 percent voted insufficient, which is five individual votes.

For the evidence of Measure 3175, this passes the criteria for evidence.

CO-CHAIR BRISS: All right, I have let the discussion of the criterion get fairly broad. I hope that will shorten the rest of our conversation. So please, if things have already been discussed, there is no need to re-discuss them.

So, anybody want to tee -- the evidence for gaps in care is on the screen in front of you. Anyone want to tee this up for us, please?

Yes, Vanita.

MEMBER PINDOLIA: I'm sorry, I'm not one of the primary reviewers. I'm not really teeing it up but when I am looking at the chart that is available for the gap, I was under the misperception that this was for new starts. So now relooking at this, how do you interpret these gaps when you don't know at what point they were

at in their treatment to then say there was a gap of discontinuation of six months if they really were at the end of the two or three year when you captured this?

DR. MATTKE: Yes, it is always an issue with long-term measurement and cross-sectional data. It is possible that we captured a few people that were on the sort of natural end of their treatment. But since this is chronic treatment, I think given that they have a two-year frame, a rolling two-year time frame during which we look at numbers I think we still have kind of the vast majority of people who are in ongoing treatment and will continue treatment.

CO-CHAIR BRISS: And so it looks to me -- so the way that I would interpret the chart, in light what Tami has said is that just on its face, based on the data presented, there is a fair amount of room for improvement if you look at the means. On its face over time, there has been some improvement over time.

And even if -- and maybe Tami gives us

some context. So even if say for the purposes of argument that a third of people actually eventually quit and are appropriately off, add a third to those means and you still get up to -- if conservatively you add another 0.33 to all those means, you are still at 50 percent or 60 percent and that is almost certainly an overestimate of the people that are appropriately quitting. So it looks to me like the measure is not specified perfectly but if they can't easily identify new starts, this still seems to suggest under treatment.

Raquel.

MEMBER MAZON JEFFERS: Wouldn't it be possible to simply add an exclusion for individuals who were on the medication for a continuous duration of I don't know a year to two years and just exclude that population from the denominator, I guess?

DR. MATTKE: Yes, it gets tricky. We had many discussions in the team how to best specify it with respect to make it cross-

sectional and capture a larger population or make it sort of new initiations and just drag those guys where you lose two-thirds of the population.

We wanted to make it encompassing. We wanted to have kind of the population on OUD pharmacotherapy and chose the cross-sectional approach over the just look after initiation.

It is in commercial claims very hard to build a long-term patient trajectories because people change health insurers, insurance quite often. So if you start putting in inclusions/exclusions that say we have to look like three years back, you decimate your sample and then you are back to the problem of having a very precise answer for a very small subset of the population.

That is, unfortunately, always the trade of the claims data like how specific are you getting versus how sensitive is your definition to what you are trying to measure.

MEMBER MARK: I mean I have done a lot of these claims studies and typically what we use

is a 30-day claim period. And you do lose some generalizability but you don't lose a lot. Thirty-day claim period just means you are excluding anyone who is on the medication for 30 days prior to the index. So you are only getting new initiators, which I think is really what you want. I mean that is the sample you want. You want to know if someone is starting this medication. They are going to stay on it for nine months. It is a pretty easy exclusion. It is done all the time. It is probably done in the antidepressant measures. I don't know if anyone knows that but it would be harmonized with that antidepressant measure if you also used the same claim period.

So, I think it could be done and it would be consistent with our other adherence measures.

DR. MATTKE: I don't dispute that you could have done it that way but the problem is if you only look at the initiators, you just lose all the people in ongoing treatment who are also

at risk with these caps. I mean it is kind of it is a philosophical difference whether you want to have a cross-sectional measure or a new initiation measure and we chose one of the two.

CO-CHAIR BRISS: So I don't see any other cards up. So maybe let's try to move to a vote on performance gap.

MS. QUINNONEZ: Voting is open for performance gap of Measure 3175. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient. Looking for a couple more votes.

Voting is now closed. For performance gap of Measure 3175, 28 percent voted high, that's five individual votes; 11 -- oh, excuse me -- 61 percent voted moderate, 11 individual votes; 6 percent voted low, one individual vote; and 6 percent voted insufficient, one individual vote.

So for performance gap of Measure 3175, this passes the gap criterion.

CO-CHAIR BRISS: Reliability. Would anybody like to tee this up for us? Raquel?

MEMBER MAZON JEFFERS: So I just had a question about the definition of the care setting being quote, unquote outpatient setting. Can you just clarify if this measure can be used in primary care settings as well?

DR. MATTKE: Yes, I think we made a bit unfortunate choice in checking boxes but we are agnostic to care settings. So it could be primary care. It could be specialty and mental health, behavioral health. But any encounter is used in care settings, et cetera.

CO-CHAIR BRISS: Mike.

MEMBER LARDIERI: Yes, thanks. Could you just explain why you stop at age 64? We're all getting a lot older and many older people are abusing substances all over the place. So why stop at 64?

CO-CHAIR BRISS: I think we did that already. So, the answer was availability of data for testing. They didn't have Medicare data.

So they have already gotten advice to broaden the pair mix and broaden the age --

MEMBER LARDIERI: I missed that. Sorry.

CO-CHAIR BRISS: Anybody else, comments for reliability? We have gotten a number of comments about specifications already. Let's try to move to a vote.

MS. QUINNONEZ: Voting is open for the reliability of Measure 3175. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

All votes are in, voting is now closed. For the reliability of Measure 3175, zero percent voted for high; 79 percent voted moderate, which is 15 individual votes; 11 percent voted low, which is 2 individual votes; 11 percent voted insufficient, which is 2 individual votes.

So for the reliability of Measure 3175, this passes the criteria.

CO-CHAIR BRISS: So and we move to validity. I'm going to tee this up. There was face validity testing where eight out ten

clinicians mostly just agreed that the results were meaningful that -- I've lost my place. Sorry. While I try to find where I am -- we have talked about so face validity testing is as we discussed. We have talked about exclusions. We have talked about meaningful differences.

And I'm open to comments from the committee. Yes, Shane.

MEMBER COLEMAN: I don't know if this is a huge deal but I just guess I want to say it out loud. Mostly like yesterday someone brought up kind of gaming the measure of sorts.

I would want to make sure that somehow this didn't incentivize systems to not offer MAT to the sickest folks who, arguably, might be the most likely to have a hard time not having a seven-day gap in the system. And in the spirit of, again, thinking through how can incidentally or accidentally do harm to folks. And I don't have an answer for how to fix that of sorts but I would worry that people -- you know if I were a system and I wanted to reach some sort of

benchmark with this measure, I might exclude the people who need it the most because I know they are the people to keep good retention rates of sorts. So I don't know. I guess I would just offer that as a caution of sorts.

DR. MATTKE: Yes, keep in mind, though that there is going to be an NCQA measure very visible measure that looks at initiation rates starting in 2018. So if you actually played that game, you would get dinged on that measure.

So we hope that together those actually capture both the ability to start patients and the ability to keep patients in treatment.

CO-CHAIR BRISS: Mike.

MEMBER TRANGLE: I really would like to hear details, if you could share them, about why most of the experts disagreed. You know?

CO-CHAIR BRISS: But eight out of ten agreed.

MEMBER TRANGLE: What?

DR. MATTKE: Eight out of ten agreed.

MEMBER TRANGLE: Oh, maybe I heard it

wrong, then.

Never mind.

CO-CHAIR BRISS: All right, I think we are wearing the committee out. Anybody else want to make a comment before we vote on validity?

It looks like we are ready to vote.

MS. QUINNONEZ: Voting is now open for validity of Measure 3175. Option 1, moderate; option 2, low; option 3, insufficient. Option 1, moderate; option 2, low; option 3, insufficient.

All votes are in. Voting is now closed. For validity of Measure 3175, 74 percent voted moderate, which is 14 individual votes; 11 percent voted low, which are 2 individual votes; 16 percent voted insufficient, which is 3 individual votes.

For the validity of Measure 3175, this passes the validity criterion.

CO-CHAIR BRISS: So on feasibility, this was calculated based on a publicly available database and the developers didn't report problems.

Anybody want to make additional comments? Yes, Tami.

MEMBER MARK: So I just want to clarify. So this will encourage providers to keep anybody on the medication, regardless of how long. I mean what is the implication of your specification is what I am trying to think through.

DR. MATTKE: Again, the implication is that you should avoid treatment gaps. You are going to see a few patients that will pretty much go off medication as a conscientious decision and carefully, not in terms of a gap but really treatment end. But other than that, we would like to see that people are in continuous treatment and don't have these interruptions.

MEMBER MARK: So by treatment gaps, do you mean treatment discontinuation ever?

DR. MATTKE: No, treatments of interrupted treatment that you have a short period during which you are not on the MAT because we do know that those periods are kind of

the highest risk people.

CO-CHAIR BRISS: Except it is true -- it would be true that the way the measure is specified that if somebody appropriately stopped these meds at four years, that would also count as a treatment gap, right?

DR. MATTKE: Correct.

CO-CHAIR BRISS: So it is one of those measures that specified in a way that you are not going to be able to get a perfect score. And you have to decide how you feel about that. That is not always a problem with the measure, in my view.

MEMBER MARK: It is not just you are always going to have a perfect score, you are going to encourage people to keep them on the medication, regardless of how long they have been on it. There is no disincentive to ever take them off because -- after the six months. But they are not -- no matter what, it is going to be measured.

CO-CHAIR BRISS: Yes, it is true. That

is probably a usability and use issue. Depending on, at least the way that I would frame that up, it is true here, as elsewhere, that sometimes high stakes measurement might be able to precipitate clinically inappropriate behavior on all kinds of measures. My residents periodically wanted to do mammograms on structure-bound people, people with six-month life expectancies. So it is possible in all kinds of measurement issues that if you are paying too much attention to the measure and not enough attention to the patient, that you can get yourself into trouble.

Yes.

MEMBER MAZON JEFFERS: So, let's say I was initiated and then I began on the medication and I stayed on the medication for six months or five months. And then I -- well, let's say six months. And then I terminated the medication. Is that called a gap or that is fine?

DR. MATTKE: That would be fine because you had 180 days of uninterrupted treatment. I think what your colleague referred to is the

situation where somebody kind of comes into our database with a two-year history of preexisting treatment.

MEMBER MAZON JEFFERS: Right.

DR. MATTKE: You don't see that history we think you are only -- we only see you have four months and we count you as insufficiently treated whereas, this may have been a conscientious decision to go off treatment.

I don't disagree that there is some measurement error in this kind of without our ability to not see the entire patient trajectory. But we wanted to err on the side of sensitivity over specificity and that is a choice. We can debate it but we thought this is the better choice, given that the performance gap is so large.

CO-CHAIR BRISS: Okay, so I think we have explored that issue in detail and it is not -- I don't consider it really a feasibility issue.

Anybody have other issues about

feasibility? Could we please vote on this criterion, please?

MS. QUINNONEZ: Voting is open for feasibility of Measure 3175. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

All votes are in. Voting is now closed. For feasibility of Measure 3175, 42 percent voted high, which is 8 individual votes; 53 percent voted moderate, 10 individual votes; 5 percent voted low, one individual vote; and zero percent voted for insufficient.

For the feasibility of Measure 3175, this measure passes this criterion.

CO-CHAIR BRISS: And usability and use. New measure so, not being currently used.

Vanita?

MEMBER PINDOLIA: Peter, I had a question. Can we ask the developer that over the next -- because it will come back for renewal in three years if it passes, right, so can we ask the developer that, over these three years, can

they track some of questions that we are having? The one that we are all concerned on is the reason for the gap. It is like a big unknown for all of us that are really trying to improve this.

I don't know if you are capable of getting that data in any sort of way but they would have the largest amount of data from lots of different plans and patient population.

And then second, to address the concern that a patient that might be ready to stop might inadvertently continue. So, if the data could be collected to say this was a new start or 50 percent were new starts, 20 percent had a prior fail before we started counting the six months, that you can tell. I understand if you are new the plan you can't but that might help during the renewal period for those questions.

CO-CHAIR BRISS: And you have gotten other suggestions as well.

MS. MUNTHALI: You can definitely ask the developer for that. And as part of our evaluation, even though it is not currently in

use, we do ask for a plan for use by your next maintenance in three years. So, I am not sure if this submission included that but that is something we would want you to get into your submission rather quickly.

CO-CHAIR BRISS: Charles.

CO-CHAIR PINCUS: What is the time frame for that?

CO-CHAIR BRISS: Charles.

MEMBER GROSS: Harold, did you get your question answered? Did you get your question answered? Oh, she's looking.

MS. MUNTHALI: Oh, now it is working. So did you have the plan ready? Tracy, does it have an end of submission? And that should be fine but with regards to the other suggestions, you are nodding, so it looks like they will be able to update the submission by the next maintenance.

MEMBER GROSS: So just to follow-up on Vanita and Tami's point I mean I think this is a useful measure but it may be the law of

unintended consequences with it being useful. And opioids being such an important topic, from a payer perspective, many state plans will see a measure like this, if endorsed by NQF and they want to attach dollars to it. And that goes to performance issues. In fact, there may be a good clinical outcome to fail with certain patients on this measure and yet from a health plan perspective, states may want to include this.

I wonder from an NQF perspective are there ways to address that sort of issue or talk about it so that new measures that need further development don't automatically slide into a contract that Anthem has to deliver on or that Rhonda has to deliver on because the state says gee, opioid, you should keep people in treatment for six months and a day, irregardless, and if you don't, there is a big penalty.

CO-CHAIR BRISS: So I think we have talked about this issue a lot already. So right now we need to vote the measure as it is currently specified and you have to decide how

important you think that issue is today.

It sounds like me it would behoove you to see if you can -- the truth is my guess is that the number of people that you pick up after two and a half years that come off appropriately before they get to 180 additional days is likely to be a relatively small number. And it sounds like, based on -- Harold I think you are recused.

CO-CHAIR PINCUS: I have a question.

CO-CHAIR BRISS: But anyway, my guess would have been that this issue is likely to be relatively small numbers of people but you are going to have to vote today knowing what we know.

And then I think it would behoove you guys to see if you can figure out how big that population actually is.

DR. MATTKE: Yes, we are happy to look into it. Now that MAT gets more common, we actually kind of a better sample to track people over time. And as you see in the national database, we only had like 40,000 in the denominator so it became really, really hard to

do long-term observations but going forward, that should be possible.

MEMBER MARK: I will just say when I looked at claims data, the adherence rate seems much higher than what I am seeing in this data set. That is another for my skepticism.

CO-CHAIR BRISS: Tami are you still --

MEMBER PINDOLIA: If I could just add comment. I think what Charles is saying is a comment I gave to Helen yesterday at the end of the meeting. When it comes to usability because now what gets endorsed with NQF has a very powerful meaning for Pay-for-Performance and incentive and provider incentives.

And I did suggest to Helen that we need to look at it a little differently now in NQF when we look at -- and I know you can't say this is endorsed for Pay-for-Performance or not but I really like the idea of being able to say that NQF at least recommends that this measure is endorsed as NQF of bringing quality. However, because of these certain caveats, can it

recommend, at least, not that anyone has to follow it, of course, but that it cannot be put into a Pay-for-Performance during that first initial phase.

CO-CHAIR BRISS: Don't we have an option for --

MS. MUNTHALI: The committee can strongly put out that statement. We can put it in the report that accompanies your recommendation that you would strongly advise application for this level. So that is something we can do.

CO-CHAIR BRISS: And don't you also have an option to endorse for further testing or something or -- what is the NQF language? There is a -- I'm getting blank looks from all the staff.

CO-CHAIR PINCUS: Let me just ask a question. In terms of overall NQF policy, so it is not just endorse or not endorse, it is endorse and you can specify things?

MS. MUNTHALI: You are giving

recommendations. They can be strong recommendations. NQF is not the committee. And part of your discussion, we would have captured this anyway in the report, but we can pull out specific recommendations for the field and for the developer.

CO-CHAIR PINCUS: So in terms of both its application, also requesting we need in X amount of time additional information?

MS. MUNTHALI: Right. So while we say that our process is use agnostic, we do know that the policy around application has changed and we are evolving. And the stakes are higher. But you did not factor this into your evaluation. This is not must pass for a reason but we want to reflect that there is some tension there. There is a strong recommendation for the failed with regards to how you see this measure being used.

MEMBER TRANGLE: Along those lines, I think we should make a strong recommendation, one, that the data be subdivided into new starts and ongoing use. And once again, I think in some

sense in my own system we are trying to track opioid use. We have defined new starts as six months and our health plan uses that and our delivery system uses that. And once you start defining just some of these fundamental blocks around which everything else is built, other things flow. But if everybody is going to have a different definition of what is a new start, how much of a clean period do you need?

And I think the systems need some definition of that around which we can kind of build other sorts of measures. Does that make sense?

So minimally, the recommendation is I don't know that I care a lot whether it is only new starts or it is everyone, as long as you can subdivide the data and we could look at both discretely.

And then two, the tobacco people have come up with a way to measure counseling in some of the measures we have already approved and endorsed. And I would encourage you to build

that in and look at some of the other methodologies.

From my perspective, these aren't enough things to sort of say I won't vote for it but it really needs to be improved and more useful.

MEMBER LARDIERI: And how is that exemption displayed? Is there like an asterisk behind the numbers so it is like visible or do people have to go into the report and read the report to get to it?

MS. MUNTHALI: They would have to go into the report. So it would not be displayed on our Quality Positioning System where you find our measures. It is our external database but they will have to go into the report. I mean we could, perhaps, bold it or something.

MEMBER GROSS: I was going to say let's vote but could the committee reconsider how they have displayed it? Nobody is going to find it in the body of the report or many people won't. You guys don't need to respond but just consider that.

MS. QUINNONEZ: Voting is now open for usability and use of Measure 3175. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient information.

All votes are in. Voting is now closed. For the usability and use of Measure 3175, 5 percent voted high, one individual vote; 58 percent voted moderate, 11 individual votes; 26 percent voted low, 5 individual votes; and 11 percent voted insufficient information, just 2 individual votes.

CO-CHAIR BRISS: We will have closing arguments before we vote on the overall suitability for endorsement that haven't already been raised.

MEMBER GROSS: So we are voting on it with the asterisk?

CO-CHAIR BRISS: Right now it sounds like the current thing for the asterisk means that the body of the report will include relevant concerns. And you have got to -- we are always voting on the state of the universe as it

currently exists.

MS. QUINNONEZ: Voting is now open for overall suitability for endorsement of Measure 3175. Option 1, yes; option 2, no.

All votes are in. Voting is now closed. For the overall suitability for endorsement of Measure 3175, 63 percent voted yes, which is 12 individual votes; 37 percent voted no, which is 7 individual votes.

For the overall suitability for endorsement of Measure 3175, this passes the recommendation criterion.

CO-CHAIR BRISS: All right, so we move quickly to the next one. So, these are sort of related to each other and so some of the issues may be the same or at least some of the issues may be similar.

So, measure developers, if you would like to tee this up for us and please try to limit yourself to not repeating issues that we already addressed in the first measure.

DR. MATTKE: Okay, so we are now turning

to alcohol use, which is a related but slightly differently specified measure just because the clinical dynamics and treatment considerations are different.

I don't want belabor the point how common and damaging alcohol use is and that medication-assisted treatment is both a supported and underused option.

Also, there are going to be two NQF-endorsed measures that are capturing the initiation so that is the same issue like with the opioid measure.

And we are looking at treatment continuity. And we do that because we know that even in the small subset that does get put on medication-assisted treatment, adherence tends to be very poor in alcohol use disorder patients. After six months, for example, in one study only about 10 percent of patients on oral drugs and about 20 percent of patients on injectables were adherent. And this is problematic because patients with longer duration and better

adherence show fewer relapses to heavy and/or frequent drinking.

So therefore, we specified that measure to focus on the continuity of pharmacotherapy, in this case defined as treatment duration of at least 180 days and sufficient adherence for the duration of treatment.

In contrast to the OUD measure, AUD pharmacotherapy is a little bit like putting patients on statins. With a linear relationship between the episodes of heavy drinking and poor health outcomes, the objective of AUD pharmacotherapy is less reduction, not necessarily complete abstinence but reduction of heavy drinking episodes.

So the definition of adherence, therefore, follows the established NQF convention of having access to the medication for at least 80 percent of treatment days, as it is done for all the other NQF-endorsed adherence measures.

Our testing results show that the measure captures substantial performance gaps

with an average pass rate of health plans around 20 percent. So some plans the rate was as close to zero percent, versus others were has high as 40 percent.

We believe that this measure will encourage health plans and providers to develop communication and education tools, as well as processes to improve treatment continuity in AUD patients and, again, we provide expert panel support for validity and usability data on reliability testing and expect that this claims-based measure will again be feasible to implement.

CO-CHAIR BRISS: Thank you. Anybody like to -- so, we will run through the criteria again, the evidence criterion.

Yes, ma'am, Rhonda.

MEMBER ROBINSON BEALE: Okay, there we go. In review of the evidence, as already stated, there is quite a bit of evidence regarding the use of pharmacotherapy for alcohol disorders.

In review of the evidence, there was a systematic review. There was also evidence reviewed in terms of quantity and consistency of the evidence and the evidence was graded. However, there was -- some of the medications that were recommended there was very little evidence to really support the length of time or having the evidence returns of the effect or the efficacy of it.

Overall, the rating by the NQF staff was one of a moderate level and primarily due to the issue with the medication itself.

MEMBER PINDOLIA: So my question, Soeren is the recommendations are all for moderate to severe alcohol for the drugs that are available that are FDA approved. And I just want to make sure I understand the denominator for this because it says just alcohol use disorder, which is a very wide range. Was the intention for this to be limited to the moderate to severe and is that the stipulation that is for the CPT codes for the entrance into the denominator?

DR. MATTKE: There is, unfortunately, no severity code in the ICD system. So we can't say moderate to severe based on diagnoses.

MEMBER PINDOLIA: I'm not good at coding but I thought I have heard people say that there is a way to differentiate that. I don't know if someone can speak up. I'm not good at coding.

Because it is really hard for me to understand and I understand that this could be similar like last time we won't ever achieve 100 percent. I get that no problem. But unlike the last one, we really might have a lot of people because there is a lot that are more on the mild. And so it would be very difficult to say as a quality metric are we really -- what really is our problem in the U.S. with this, how bad is it, and what are we trying to improve it to if we really don't know how much of a noise we have down there?

DR. MATTKE: I mean the advantage in this case is, Vanita, that we don't have the initiation measure. So we are not judging

whether or not the decision to put somebody in treatment was appropriate. We basically say okay, we respect the clinical decision, which we cannot second guess, based on the data that we have at hand. But if you put somebody on these drugs, you had better make sure that they take them for at least 180 days and sufficiently adhere.

MEMBER PINDOLIA: That helped clarify. Thank you, Soeren. I didn't think of it from that perspective.

MEMBER MARK: Yes, the evidence cited is from a Jonas paper and also from the Department of VA and DoD. And if you look at the Jonas paper, it actually doesn't say that adherence, lack of adherence leads to relapse. In fact, it points out that many people discontinued before they even relapse. So they can't actually -- what it actually found was people began to drink before they even discontinued. So, they didn't find that once people discontinued it increased relapse. They found that people started drinking

even before they stopped the medication. So they didn't find -- I didn't read the paper as supporting that.

And then when you read the VA and DoD guideline, they don't actually talk about the need for adherence. The guideline just says that medication must be offered to anybody who has an AUD.

So actually, I am very supportive of the continuity of care issue for OUD. I think the evidence is pretty strong for nine months. I think it is very different when you get to AUD. The evidence is really not there to support nine months.

And then just finally on the evidence, if you look at the largest clinical trial, the combine, they find that cognitive-behavioral therapy does just as well as medication in reducing AUD. So, people are going to do very well in cognitive-behavioral therapy in contrast to I think OUD, where the evidence is much more stronger that yes, that you need medication. I

mean relapse can be fatal.

DR. MATTKE: So I think that there were several questions embedded. Unfortunately, no guideline ever says if you recommend treatment we also recommend that the patient is adherent to treatment because our inherent bias is that yes, if it is a recommendation the patient trajectory also follow it if it is prescribed. So I have yet to see a guideline that says adherence matters.

Yes, I agree that there is a greater number of patients that do not need drugs. But as I said in my response to Vanita, we respect the prescriber's decision to put a patient on MAT and just then follow whether or not adherence is there.

I think I lost the third question. Yes, none of these drugs actually are getting people to be completely abstinent. It tends to be a risk reduction strategy in which as long as they take these drugs reasonably regularly, they tend to have fewer drinking episodes.

And since there is very good evidence that the harms from alcohol are a linear function of sort of your exposure to heavy drinking, that is the treatment objective. And that is slightly different from the treatment objective in opioids, which is why we specified that component a bit differently.

MEMBER MARK: So I just forgot one other thing. Yes, I think the counterargument is that if there would be no negative, that would be fine. Just give everybody the drugs. But these are associated with some serious side effects and so that is the -- you have to balance the harm against the positive.

DR. MATTKE: But again, that is the prescriber's and patient's decision, which we are not measuring. We are saying if you make that decision jointly, you should stick to the treatment plan for a specified duration of time.

MEMBER MARK: Regardless of whether they have side effects or not after they start the medication.

DR. MATTKE: Yes, you could stop the treatment.

MEMBER ZUN: I'll be brief on my soapbox today but I figured you know I needed to do it once a day.

So, if we know there are other treatment modalities that are non-pharmacologic that work as well as the pharmacologic, why don't we have a measure that looks at those options, rather than the one, the easiest one out, which is the one that we know about that is measurable?

I do believe there are billing codes for counseling, and CBT, and all those other things. So, why are we just focused on one treatment modality when that is not in the patient's best -- may not be in the patient's best interest?

I'm done. Thank you.

CO-CHAIR BRISS: As I used to tell my kids, I think we need to talk about the definition of need. Lisa.

DR. WATKINS: This is Kate Watkins. I think that is a really good point that that is

not what this measure does but I do believe that patients should be given choice. And this measure deals with one aspect of treatment, which is pharmacotherapy.

MEMBER SHEA: So, I know these are complicated issues but unlike the opioid use disorder, I was just wondering what the evidence was for continuing medication for alcohol use disorder for that amount of time is one question.

Two, in my clinical experience, you are often switched. People switch the medications because they have an issue or whatnot. So, they would look like they have dropped off the medication but be on another one. So that was my other question. How do you account for that and how do you account for sometimes people are on more than one of these at a time?

And then my last question was there is a couple of them that are used for other reasons, as well. They are not even necessarily FDA-approved for alcohol use disorder. So how would you know if a person stopped the gabapentin or

the topiramate if it was the alcohol use disorder that they were on it for that in the first place?

DR. MATTKE: Good questions and those we love.

So the treatment duration we justified by the fact that the approval trials for the drugs that have an FDA label typically ran for about 180 days. So, we can basically say we don't really have good evidence that shorter treatment durations work and so we take that as a minimum limit.

The switches we would capture. We are not forcing you to be on one drug. So if you switch somebody from naltrexone to gabapentin, we just count the number of days in treatment and you would still be in adherence. The same for more than one, if they overlap we don't double count the days.

Other indications, Kate, you can probably speak to that better. That was the trickiest thing to deal with, especially with gabapentin because it has such a broad range of

indications and is not usually labeled for alcohol but recommended for it. But Kate, you want to explain on that gabapentin?

DR. WATKINS: We gave you a pass, essentially. If you had a diagnosis of an alcohol use disorder and were being prescribed gabapentin, that you got a pass for being on it.

DR. MATTKE: Yes.

MEMBER TRANGLE: I mean the reverse of that would be if you are on it for diabetic peripheral neuropathy and you stop it for that, you would get dinged.

DR. MATTKE: Yes, I mean that was one of the choices we had to make. We give you your credits or are we taking it out because it is not specific to AUD. I mean you be the judge whether we make the better choice.

CO-CHAIR BRISS: And Soeren, I am going to take off my chair hat for a second and make a couple of comments.

I sort of share many of Tami's concerns on this one. I think the evidence for

persistence is of the same quality for this as it is for opioid use disorder. And when you actually look at the effect sizes in the studies, they are modest I would generally say. And so whether this, as it is currently specified really -- and then there is all of that makes the issue of leaving out talk therapies, which might be preferable sort of relatively more important and so I wonder whether all that stuff changes the thinking on importance to measure and report.

Raquel.

MEMBER MAZON JEFFERS: So people have sort of alluded to this in their comments but I just want to highlight it. Two of the medications that you have listed are not approved by the FDA for the use of alcohol use disorders. That is a really -- I mean I am normally not on the conservative side of these kind of conversations but that is a really important consideration for a committee like this to approve as a measure a medication that hasn't been approved by the FDA for this purpose.

DR. MATTKE: Yes, it was a tricky choice because the guidelines explicitly say we recommend using it, even though there is no label. And given that some of these drugs have been around forever, there is no manufacturer who will do a trial to get a label because there is no commercial interest involved. And that, unfortunately, is occasionally the case with these old generic drugs that nobody cares anymore and then the guidelines go based on off-label use, even though the FDA has never endorsed it.

MEMBER SPERLING: So as the patient advocate in the room I feel compelled to speak up here. This is, I mean it strikes me as dangerous territory for NQF to be implicitly endorsing off-label use. And if it is going to be done, I strongly recommend that it at least be tethered to peer-reviewed treatment guidelines from respected medical society like the American Society of Addiction Medicine or be part of peer reviewed literature that has something to back it up.

Because the FDA label isn't the only thing. Because in psychiatry I can tell you have lots of off-label treatments but they are backed up by strong evidence, endorsed by APA treatment guidelines. And I know the American Society for Addiction Medicine does a lot of this.

But let's just make sure we are tethered to the science, that's all, from the patient safety perspective.

CO-CHAIR BRISS: You may or may not agree with the guideline but they list the guideline on which their recommendations were based. So, you may or may not consider the VA/DoD to be the most appropriate guideline developer but there is a guideline that they have tried to tether to.

So, David, you were next I think.

MEMBER EINZIG: So just qualifying this with this is not my patient population but for those who work with this population, a question for the room, is this an accepted standard of care that people are using?

And then my second question is if the medications aren't helping and you stop and after a couple of months, why continue something that is not working?

DR. MATTKE: Well, the last is easy. If you stop after 180 days, you are fine because you have tried it for at least that period. So you wouldn't get dinged.

The first, of course, it is a recommended, a guideline-recommended treatment option and it is only of the options, as Kate says but we are not looking at whether it was appropriate to put the patient on that option. That is the decision of the patient and the prescriber. We are looking at if the patient is on that option, are they getting treatment for evidence-based period.

CO-CHAIR BRISS: Yes, sorry. Any of the relevant clinicians in the room want to talk about standard of care issues?

MEMBER COLEMAN: I mean I would just say that to me it is tethered around that six month,

180-day period. I don't know that that is the standard of care, per se, identify it as yes, it is going to be way more effective if you can reach that point. I think people have kind of already mentioned the difficulties with what that evidence base may be.

Certainly, the medications are pretty standard of care. Those are the exact medications I would have listed off as potential options. But again, this measure kind of comes in after that has already been said. Like they mentioned, it is about continuity I guess of the medication or the time without gaps, things like that.

So I think we are questioning the evidence base for this for exactly what they are measuring is part of the problem.

MEMBER TRANGLE: Yes, and it is maybe not binary. You know the evidence for gabapentin and Topamax would not be the same as acamprosate or naltrexone or Antabuse.

So, I wouldn't answer totally thumbs up

or thumbs down for the evidence. It varies.

You know and sometimes we use stuff when somebody is relapsing so much. We have tried everything else, so what the hell? You know you can grade evidence but it is part of the standard of care then.

MEMBER KELLEHER: Yes, I am not a prescriber but Mike's variance in being in agencies where there is prescription mostly for a dual diagnosis, that they may not be the first line of defense but if other things are not working for whatever reason, it is common, I think, to keep going to try to find something that works.

My question is if the data shows that I am one for two months and then all of a sudden that is discontinued and now I am on another on the list, that would count as a continuation or no?

DR. MATTKE: Correct, yes.

MEMBER KELLEHER: It would still be a continuation?

DR. MATTKE: Yes.

MEMBER KELLEHER: So if I keep changing because of one reason or another but it is on the list, that would be considered still continuity.

DR. MATTKE: Correct.

MEMBER KELLEHER: Correct?

DR. MATTKE: Yes.

MEMBER KELLEHER: All right.

CO-CHAIR BRISS: Are we ready to vote on the evidence to support the -- Tami, first.

MEMBER MARK: So there was this randomized trial that show people who discontinue buprenorphine, they relapse. Is there a similar randomized trial that shows people who discontinue acamprosate or naltrexone they relapse?

DR. MATTKE: I'm not sure that I understand the question.

MEMBER MARK: So I am saying the evidence that it is important to take the Suboxone for 180 days comes from a randomized trial that shows that when people discontinued,

they had a very high rate of relapse. And I am wondering if there is comparable evidence on the AUD side to show that if people discontinue -- from a randomized trial, that if people discontinue before 180 days on naltrexone, there is a huge -- what is the rate of relapse?

DR. MATTKE: Yes, so the approval trials for these drugs typically ran for about six months. The endpoint is typically not relapse because in AUD the treatment objective is sort of risk reduction, as opposed to abstinence. So you want to avoid as many heavy drinking days as possible.

So most drugs were approved based on an endpoint of fewer heavy drinking days, not based on the endpoint of complete abstinence of alcohol. But yes, there is randomized --

MEMBER MARK: Even if you weren't calling heavy drinking relapse, is there a similar study that shows that when people discontinue there is a huge risk of returning to heavy drinking? That is the evidence that

supported -- if you want to explain it.

DR. MATTKE: The evidence is there that as long as you are on the drug, you have fewer drinking days. So that is basically showing if you are off the drug you have more drinking days.

CO-CHAIR BRISS: Michael.

MEMBER TRANGLE: I can comment a little bit. I mean I think the evidence is kind of low to moderate level evidence and it really is the way you talk about it.

The relapses that are going to occur will likely be not as intense drinking during the relapse and maybe a little bit shorter but not versus no relapses.

Does that answer your question kind of?

MEMBER MARK: Yes, just some perspective. What is your perspective on the risk of these medications? Like I have heard some clinicians say you know naltrexone, you have to override if you have surgery because it is a -- or you know the side effect profile seems pretty safe.

MEMBER TRANGLE: I don't see them as super dangerous. I mean you know in some sense Antabuse has been around forever and that can injure your liver but use it if it works. I don't know. You guys should comment on that, too.

MEMBER COLEMAN: I'm happy to bring -- should we bring it up later though in usability or something like that or later? I don't know if the evidence-based is the right time to --

CO-CHAIR BRISS: I'm okay with calling this evidence.

MEMBER COLEMAN: I mean so I am mostly in favor of using these medications so I don't want this to come across the wrong way but naltrexone certainly has challenges. It is actually more on the Suboxone side than it is on the alcohol side because usually naltrexone you can just stop prior to surgery. It's no big deal. With an opioid use disorder because folks go into withdrawal and that is the biggest thing you are trying to avoid. It becomes a much

bigger deal because folks can't just go off of it. And I do have campus-wide hospital procedures that we had to develop that are not perfect that none of the medical docs like dealing with. So it is Suboxone actually more than this maybe is an issue.

CO-CHAIR BRISS: So I think -- I wonder if we might be ready to take a vote.

So again, this is on evidence that this measure, as specified, is going to result in people being better off.

MS. QUINNONEZ: We are now voting on Measure 3172, Continuity of Pharmacotherapy for Alcohol Use Disorder. Voting is now open for evidence. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient information.

All votes are in. Voting is now closed. For the evidence of Measure 3172, zero percent voted high, zero votes; 37 percent voted moderate, 7 votes; 47 percent voted low, 9 individual votes; and 16 percent voted for

insufficient, which is 3 individual votes.

So for evidence of measure 3172, this does not pass the criteria for evidence.

CO-CHAIR BRISS: So generally speaking, we have given -- I thought the feedback was pretty clear, actually but generally speaking, we have given the developers an opportunity to ask any clarifying questions about anything further you would like to hear from the committee that led to this conclusion.

DR. MATTKE: No. I mean we do know that the evidence for AUD is weaker than for OUD, it is just a much less researched issue than OUD. We will just try to keep at it and see what we can do moving forward.

CO-CHAIR BRISS: Thank you. And presumably, some of the issues, as you are reworking both of these, or if you decide to rework both of these, many of the same issues that applied to the OUD would likely apply here. So, the breadth of the population and things and counseling.

Yes?

MS. JOHNSON: Just a quick question from the staff side. Since we have to write this up, we want to make sure we understand exactly why you landed where you did.

So three things that I think I heard was the list of meds, the 180 days, and maybe a little bit of VA/DoD guidelines versus others that might exist. Did I have that right? Are those the main issues that you guys had? Would somebody just help us understand?

CO-CHAIR BRISS: I might have prioritized slightly differently. So what I heard was that the biggest issue was probably the evidence for 180 days. And then secondarily there may have been issues with -- there was hung jury, as I heard it on the meds. There were several people that thought that those were the appropriate ones and some people that were worried about the FDA approvals, about which I didn't hear a single conclusion.

MEMBER MAZON JEFFERS: So I also think

there was one other recommendation on around the use of the efficacy of CBT being on par with pharmacotherapy for alcohol use disorder.

So unlike for opiate use disorder, where there is overwhelming evidence that medication-assisted treatment is more effective than counseling alone, I think this is not the case for the alcohol use disorder medication. There is evidence to suggest that the use of medications to support recovery from alcohol use disorder is not that much more effective than CBT alone or counseling alone.

MS. JOHNSON: Okay. And to that last point, I just do want to make sure that you guys didn't vote low or insufficient on this measure as put before you because it wasn't constructed in the way that you would like to have seen it.

So in other words, your vote would have been based on the evidence that was presented for the medications. And again, you guys know I am not an addictions person. So I am struggling here.

CO-CHAIR BRISS: Mike or Shane.

MEMBER TRANGLE: For me it really had to do with the strength of the evidence of these medications working, how well they work, and what impact they have and the quality of the evidence to support that.

CO-CHAIR BRISS: And Shane.

MEMBER COLEMAN: I guess I am just asking is what you are saying is that if we are interested or if someone were interested in not supporting this on the basis of the CBT or other that it would have been maybe a different category of sorts that we would have voted down on, other than evidence. Is that what you are suggesting, like usability or something like that?

Because I do think -- and the reason why I bring this up for clarity is because I do think you are going to come up against a tension both with opioid and alcohol but certainly of a non-pharmacologic versus pharmacologic. And I am guessing that there are some folks that I guess

without the inclusion of the non-pharmacologic don't feel comfortable supporting because it may drive care and it may emphasize pharmacologic without the non-pharmacologic.

So I guess another way to say that is where would you have voted it down for that reason of sorts or something if it wasn't in evidence.

CO-CHAIR BRISS: And my friendly amendment to that might have been that I feel like there was plenty of concern about the evidence about the 180 day stuff to result in a no vote.

I also think that in addition to that, it does appear that this could have the effect of supporting medication use that is not necessarily fully supportable on the evidence over other options that might be better options. And so if you don't have a place to currently put that, that kind of thing as we are thinking more about portfolios of measures, we ought to have a place to put that.

Anybody else? Mike.

MEMBER TRANGLE: My comment really isn't for you guys but in some sense this seems like we get a lot of measures that are about little slices of kind of treatment things that we might be doing about something and we have got no measure for some of the basics.

You know it is like if I think about my system, the things we are struggling with is we are trying to come up with an opioid measure to figure out how many new prescriptions we are having. When do they turn to chronic prescriptions, which we have defined as three scripts in six months of opioids? You know and why are we -- what are the categories that they are on for? And what is the number of pills in a morphine equivalence?

We are just trying to get our handles on the basics of sort for how many opioids we are using in a system and that is like a fundamental building block. And we are talking about a little slice here and a little slice there that

is easy to measure but not fundamental to how are we doing.

CO-CHAIR BRISS: So we are going to have a lot of time this afternoon to talk about the whole portfolio and I want to be respectful of our RAND colleagues' time. So if you have additional comments that you want to particularly make while our RAND colleagues are here about this measure, let's do that. I need to then move us quickly to public comment.

And if we want to talk about the whole portfolio and how things fit together, I would rather move that conversation to this afternoon, if we can.

CO-CHAIR PINCUS: I just have one comment that actually is not related to this measure per se but because Kate -- are you still on?

DR. WATKINS: Yes, I am still on.

CO-CHAIR PINCUS: Yes, I think it would be useful to talk about sort of our experience in looking at some of the VA data in thinking about

substance abuse measures more generally and the problem of in the absence of having a screening program, how one looks at continuity because of selective choices that people make in terms of having an initial visit. I think that is counted.

So, do you want to comment on that a little bit? Do you know what I am talking about?

DR. WATKINS: No, I'm not sure I do.

CO-CHAIR PINCUS: Okay, so I guess I have to make my own point and Kate you can chime in.

So, Kate and I were involved in a study looking at the VA in terms of the quality of care for mental health and substance use conditions within the VA. And among the findings we had in comparing the VA performance on administratively based claims measures to a private sector data base, we found that the VA, by and large, did somewhere around 5 to 15 points better on most measures, with the exception of substance use measures of engagement and initiation engagement.

And in part, it was based upon a huge difference in prevalence between the two populations. In the private sector there was about a one and a half percent of an initial encounter; whereas, in the VA, because they were doing screening and maybe also in terms of the prevalence there, it was a much higher prevalence, actually an order of magnitude higher prevalence.

And so our hypothesis was that part of the reason why you were seeing this difference was because by screening, you were getting a less motivated -- it goes to your point, Mike -- a less motivated group as compared to people that actually in the private sector were coming in because they wanted treatment and that is how they got captured.

And so and Kate you may want to comment more on this but I thought that just for our discussion later, we should think about how to think about sort of having more of creating a common baseline around these sorts of measures.

CO-CHAIR BRISS: So, I actually think we have been competing a little on how fast we can get through measures as the two chairs and I consider this to be an unseemly use of the four corners to run up my time.

So I want to -- anybody else got specific stuff for our RAND colleagues before we let them go?

Shane, do you have something specific that you want RAND to hear? No.

So, thank you.

And now I think we need to do public comment before we break for lunch.

OPERATOR: Okay, at this time if you would like to make a comment, please press star, then the number 1.

CO-CHAIR BRISS: And I'm sorry. Public comment can also come from people that are here in the room.

OPERATOR: There are no public comments from the phone line.

CO-CHAIR BRISS: So we actually have a

long lunch break scheduled. We are off until 12:45.

CO-CHAIR PINCUS: Do you want to make it shorter so --

CO-CHAIR BRISS: Let's reconvene at 12:30. Is that okay? Good.

(Whereupon, the above-entitled matter went off the record at 12:10 p.m. and resumed at 12:38 p.m.)

CO-CHAIR PINCUS: So Erin, do you want to --

MS. O'ROURKE: Absolutely, thanks, Harold.

So I should probably start by apologizing to Harold. He's heard this presentation before. So I'd welcome any reflections you have, as we also shared this work with the MAP Coordinating Committee, to really think about how attribution impacts things when the rubber hits the road and measures are put into use.

CO-CHAIR PINCUS: And I just may -- I

want to point out that I think this issue about attribution and accountability is especially relevant -- and Erin and I were just talking about this -- for this committee.

Because inevitably, there is an interface between behavioral health and general healthcare that sort of mushes up the whole issue of accountability and enables finger pointing about who screwed up and who's responsible, and whether it's at the provider level or at the healthcare organization level or at the plan and carve out level, these kind of things come up.

Also it's an area where, because of many of the sort of long-term aspects of these conditions involve sociodemographic mediators of health. And it's unclear who's responsible for dealing with some of those issues.

So this is something that I think is especially important.

MS. O'ROURKE: Thank you.

CO-CHAIR BRISS: I'm sorry, one more point that needs to be made about this is that

attribution gets harder as we're all trying to move toward outcomes measures. And the farther you get toward outcomes, the harder the attribution gets.

And we lived through some of that on the population tobacco measure yesterday. So Erin's going to solve all those problems for us in the next 30 minutes.

MS. O'ROURKE: Well, with that introduction, so, I'm Erin O'Rourke. I'm one of the senior directors here at NQF supporting the work of our attribution expert panel.

So we've continued to see through recent legislation such as IMPACT and MACRA this focus on value-based purchasing as a way to drive improvements in quality and costs.

However, implementing these new payment models means we need to know who can be held responsible for the results of the quality and efficiency measures that are used to judge performance through these programs.

And as Peter was saying, there's also

simultaneously this increasing push to measure quality through outcomes, and that makes it even more challenging to know we can hold accountable.

So attribution is the process that tries to determine who can be held responsible. It's the methodology used to assign patients and their outcomes to clinicians or providers.

And attribution models attempt to identify a patient/provider relationship that can be used to establish accountability for quality and costs.

So as we continue to move away from fee-for-service payment to alternative payment models, we need to better understand how patient outcomes and costs can be accurately attributed in a system that's increasingly built on shared accountability, particularly, as Harold was saying, in fields like behavioral health where it really is a team effort.

So taking all of this account, NQF launched a project to provide guidance on attribution issues.

Specifically through this project, we aim to identify key challenges to attribution, develop a set of guiding principles, identify the elements of an attribution model, and explore the strengths and weaknesses of various approaches.

And then finally, identify recommendations for developing, selecting and implementing an attribution model.

So we brought together a multi-stakeholder group to really try to advance the science behind this area of measurement.

You can see it was co-chaired by Ateev Mehrotra of Harvard and Carol Raphael of Manatt Health Services.

We included stakeholders across the care continuum. We had developers, clinicians, hospital representatives, methodologists, consumers, purchasers, payers, and suppliers.

So to inform the committee's work, we started by commissioning a white paper from a group of authors out of the University of Michigan and the University of Pennsylvania.

They performed an environmental scan of the attribution models currently in use. They found huge variation in how models are characterized and what the different elements included in each model are.

So I'm guessing this won't come as a surprise for many of you that deal with this every day, but I was at least pretty taken aback by the scan turning up over 160 models that are currently in use or proposed for use.

And the vast majority of these use retrospective attribution and attribute it a single provider, usually a physician.

So the commission paper included some pretty interesting findings. The authors noted that, currently, best practices for attribution have not been determined.

Rather, new models are just built off of existing approaches, so we just keep using what's been previously used without really taking the time to study the trade-offs of the different approaches to an attribution model.

And really, the authors called that there needs to be more transparency here, and we need to explore the strength and weaknesses of different approaches.

There's also currently no standard definition for an attribution model, and this makes it really challenging to compare them.

We have a limited ability right now to compare cross models to evaluate what works best. And it's really critical to developing the evidence base that we get to a greater point of standardization so that we can start to make determinations of best practices.

So the committee wanted to tackle some key challenges through this work.

First, they noted that greater standardization is needed among attribution models so that we can start to make those comparisons between models and allow best practices to emerge.

The committee found there's little consistency across models, but there's quite a

bit of evidence noting that changing the attribution rules can dramatically alter results.

So in turn, this can really change how a clinician or a provider might score on a performance measure or in an accountability program.

There is also a lack of transparency on how a patient and their outcomes are attributed. This means there's often no way for a clinician or a provider to appeal the results of an attribution that may wrongly assign accountability.

So to start to address these challenges, the committee came up with a number of products through this work. The aim of these products is to allow for this greater standardization of attribution models, to increase transparency around attribution, and to hopefully get greater stakeholder buy-in so that we can allow for better evaluation of attribution models in the future and start to lay the groundwork to develop a more robust evidence base.

As a first to addressing these attribution challenges, the committee agreed on a set of core principles to ground its recommendations.

These principles acknowledge the complex multidimensional challenges to implementing attribution models, as the models can change depending on their purpose and the data that's available.

The committee grounded its work in the goals of the National Quality Strategy. So better care; healthier people, healthier communities; and smarter spending.

And the committee recognized that attribution plays a critical role to advance these goals and to continue to drive improvement across the healthcare system.

So attribution can refer to both the attribution of patients to a clinician, group of clinicians, or facility for accountability purposes, as well as the attribution of results of a performance measure such as an outcome or

resource utilization to a clinician or facility.

The committee really highlighted that, currently, there's no gold standard for designing or selecting an attribution model. It's therefore important to understand the goals of attribution for each specific case when you're assessing potential models to apply.

Some key criteria to consider when selecting a model are actionability, accuracy, fairness and transparency.

This is particularly important as the application of an attribution approach can really significantly influence the reliability, validity and results of a measure.

Moreover, attribution can significantly affect the size of a population for whom a provider is assigned responsibility as well as determine their performance under value-based purchasing programs.

So on this slide, you can see the committee's guiding principles. The committee recognized the importance of a trusted

patient/provider relationship and the need to continue to enhance patient centeredness and coordination of care when developing attribution models.

Again, attribution models are a set of rules used to logically assign accountability for a patient's care and to help drive improvement.

The term provider is defined broadly here to include individual clinicians, clinician groups, hospitals, other facilities like skilled nursing facilities, system levels, ACOs, et cetera. So again, provider is kind of default term.

However, for the purposes of measurement and payment, it can be challenging to determine a patient/provider relationship, particularly for outcomes where multiple providers may share responsibility.

So the committee recognized the current tension between a desire for clarity about a model's fit for purpose and the state of the science related to attribution.

There's a desire in the field for rules to clarify which model to use in a given circumstance, but the committee really felt that, right now, there is not evidence to support the development of such rules.

As noted above, a significant finding of the paper was the current lack of standard definition included in attribution model and how this lack of standardization across models really limits the ability evaluate the effectiveness of different approaches.

So as an important first step to evaluating attribution models, it's necessary to determine what elements need to be specified.

And the Attribution Model Selection Guide that we'll cover in a few slides is intended to aid developers, evaluation committees, and program implementers on the necessary elements.

Apologies, I have a bit of a cold, so my voice is a little froggy.

So the guide is really meant to enable

stakeholders to have a structured dialogue about the various models and the decisions that should be made when implementing a model.

The Attribution Model Selection Guide represents the minimum elements that should be shared with accountable entities.

The detail of an attribution model and the choices made in developing a model should be transparent to patients, accountable entities, and other stakeholders.

An attribution model must be well tested, defined, and precisely specified with adequate testing so that it can be implemented consistently. The Attribution Model Selection Guide includes a series of key questions to answer in development and selection of the model.

Again, it's designed to improve standardization across models and to increase the ability to evaluate models in the future.

So on this slide, it's a bit hard to read, but we did want to share with you the Attribution Model Selection Guide. Again, I

won't belabor this, but it takes the user through a series of key questions and asks them to at least think about these elements when they're developing their model.

Again, there wasn't really enough evidence to come up with set answers to each. If you look at the report, the committee provides some guidance on some of the pros and cons of different approaches and some key considerations that you might to take into account as you're developing your model.

But at least we wanted to get people thinking about these issues and perhaps thoughtfully considering the trade-offs when they are developing an attribution model.

So the committee's recommendations build on the principles and the Attribution Model Selection Guide. The recommendations are intended to apply broadly to developing, selecting, and implementing an attribution model in the context of both public and private sector accountability programs.

Through these recommendations, the committee attempted to recognize the current state of the science and to consider what's achievable now and what is the ideal state for the future.

The committee stressed the importance of aspirational and actionable recommendations to develop or to continue to drive the science of attribution.

So the committee's first recommendation is to use the Attribution Model Selection Guide to evaluate the different factors that should be considered in a choice of an attribution model.

The committee emphasized there is currently no gold standard for attribution, and a different approach might be more or less appropriate in a given situation. The choice of a model should depend on the context of its use and should be supported by evidence.

It's also crucial to be transparent about the potential trade-offs between the accountability mechanism, the potential for

improvement, the degree of control a provider may have over an outcome, and the scientific properties of the measure.

The next recommendation is that attribution models should be tested. Attribution models of quality initiative programs should be subject to some degree of testing for goodness of fit, scientific rigor, and unintended consequences.

While the degree of testing may vary based on the stakes of the program, attribution models would be improved by rigorous, scientific testing and then making the results of that testing public.

Again, the committee stressed that sometimes pilot testing may be available under certain circumstances such as private reporting. This could help to generate the data to understand what the model is achieving and if it's doing what it's intended to do. And then if so, that model could be used for high stakes applications such as payment or public reporting.

When used in a mandatory accountability program, models should be subject to testing that demonstrates adequate sample size, appropriate outlier exclusion and/or risk adjustment to allow fair comparisons among attributed entities and sufficiently accurate data source to support the model.

The committee recognized that data for an attribution model can really vary. It could include claims, electronic health records, clinician attestation, or patient attestation.

So the next recommendation is attribution models should be subject to multi-stakeholder review. Given the current lack of evidence on a gold standard for attribution models, a stakeholder's perspective can really influence what is best and what model they feel may be most appropriate.

Again, the committee emphasized that this is an opportunity to bring together stakeholders across the continuum to review the models and determine if it's really the best fit

for its intended purpose.

The committee emphasized that models should attribute care to entities who can actually influence the care and the outcomes, recognizing that, currently, models can unfairly assign results to a provider that has little control over the patient outcomes.

So for a model to be fair and meaningful, an accountable entity must be able to at least influence the outcomes for which it's being held accountable, either directly or through collaboration with others. Again, trying to recognize that we do continue to move to shared accountability, but a lot of these models tie everything back to one provider.

So as care is increasingly delivered through teams and as facilities become more integrated, a model should reflect what accountable entities are able to do to influence rather than directly control.

And then, this was one that the committee really wished to highlight, so this is

their final recommendation.

A set of minimum criteria for models used in mandatory public reporting or payment programs. Again, the goal here was really to try to improve the current state and get to a greater sense of buy-in and fairness from what patients are being attributed to a set provider and how their results of that patient's care are being used to determine what's reported to the public and how providers are being paid.

So I think with that, I can take questions or open for any discussion.

CO-CHAIR PINCUS: So questions, comments that people have?

Raquel?

MEMBER MAZON JEFFERS: I was just wondering if you could elaborate a little bit about how you see this issue of attribution filtering into the kinds of measure review processes that we're currently engaged in?

MS. O'ROURKE: Absolutely. So that was something the committee recognized that we need

to start considering a little bit more in NQF's review processes, both for endorsement and selection.

It's kind of an underlying tide, if you will, but it's not something that's explicitly on the table right now.

So the committee didn't really have set recommendations about where we could put this in the NQF criteria. They explored, perhaps, once there is a better evidence base, it's something we could look at under the validity subcriterion to, you know, start to at least make developers explain how the results of a model might be attributed and allow the evaluation committees to determine if they think that's appropriate or not.

So again, no set answer at this time, but really, the call was to we need to build this evidence base so that we can get there in the future.

CO-CHAIR PINCUS: Rhonda? And then Peter, Brooke and Vanita.

MEMBER ROBINSON BEALE: Just a couple of quick questions. You had quite a few experts on the panel. I'm just curious whether or not, because I don't know the names, whether or not CMS was represented there? Organizations like Mehlman, Truven, these are organizations that are deeply in doing attribution modeling.

MS. O'ROURKE: Absolutely. So, we did have a liaison with CMMI, actually. So, not out of the group we usually work with at CMS, but out of the Innovation Center.

We also had some group representatives, not from Mehlman or Truven, but we did have developers from NCQA, Yale CORE, RTI. Let me see, we had some experts out of groups like Advisory Board.

So we tried to get the -- a representative from some of the different groups who are involved. But again, obviously, could not be complete given we had only 25 spots on our committee.

MEMBER ROBINSON BEALE: And the other

question, are your slides going to be made available to this committee?

MS. O'ROURKE: Absolutely.

CO-CHAIR PINCUS: Yes, I believe the report is available.

MS. O'ROURKE: Yes, the slides and the full report are publicly available. Yes, it's posted on our website. It was funded by CMS, so this is all in the public domain, and we welcome you to share it with anyone who might be interested.

I can get the link to -- Desi and Kirsten to share with you all after the meeting.

CO-CHAIR PINCUS: Peter?

CO-CHAIR BRISS: So Erin, thank you for that. So I'm a notorious heretic; I'm going to be a notorious heretic again.

I actually think we're asking the wrong question for the most part. So it's -- the truth is, medicine is almost always a team sport these days. And these things are, generally speaking, trying to attribute the structure, process, or

outcomes to individual physicians.

And so we're preparing for an in case William Osler comes around again. You know? But I don't think he's going to be back.

And so I know there's a big push from at least -- advocate consumers would really like to have kind of physician-level scores. But I don't think it's reasonable.

And actually, if you -- I'm afraid it interferes with setting up systems that encourage team-based care in the way that we should be encouraging team-based care.

So I actually think we're -- I actually think this is, in some ways, a move backwards. And if we want to be talking about attribution, it may be, outside of a few procedural specialties, we ought never to be talking about individual physicians ever.

MS. O'ROURKE: Absolutely. And I think that was some of the issues the committee really grappled with, that you want to encourage the move to team-based care. But like you were

saying, recognize that consumers are trying to shop around for an individual physician and want that level of information available.

I think, in particular, they also struggled with, as we are moving out of fee-for-service and towards population-based payment, how do you do this?

So I think, again, it was trying to recognize where we are currently, and particularly in this one, if you are going to attribute to an individual physician, set some guidelines around that to try to increase fairness and accuracy.

But absolutely, I think the committee was with you on the need to recognize that medicine is a team-based sport, and we need to find ways to hold the team accountable.

I think, in particular, someone used a baseball metaphor of the teams win or loses, but it's perhaps to the team to determine which player was MVP.

MEMBER PARISH: Yes, so very concrete

question. Is the report -- I know there's no gold standard, but did it come up with a top five or anything like that? And does it have the top five with sort of recognition of the weaknesses of it?

MS. O'ROURKE: So they do go into the pros and cons of different approaches, recognizing there is a number of ways to do this, in particular, a push to move to patient attestation, and that other groups such as the Health Care Learning and Action Network said that perhaps that is the gold standard.

Our committee noted there are some limitations to that approach. They explore the different pros and cons in the paper more, so explaining.

MEMBER PARISH: But, that's all in the public domain and on the --

MS. O'ROURKE: All in the public domain.

MEMBER PARISH: Okay.

MS. O'ROURKE: We can share that with you if there's a particular approach you wanted

to see more what they said.

MEMBER PINDOLIA: So more of a comment, I think, and then, maybe a question at the end. I understand where Peter's coming from. But to be honest, it's really difficult on the provider end when we have one pair saying this -- we're going to attribute these patients to you, and you're in our PPO plan, so there's no assigned PCP. And then, there's another with HMO, which is very clean: you're the assigned PCP.

So I am really happy that there's been maybe some kind of standardization to kind of help guide that a little bit more for our providers. And also from the payers perspective when we do have our PPO.

But for behavior health in particular, this has become a great challenge for especially our opioid shoppers, and they just use ER, and I don't know who to attribute. And so I can't get a provider contract for those patients to kind of lock them in to be able to limit their opioid use in any sort of way.

So I think it's actually even more needed -- needed more for us to help manage our behavioral health, if we can try to figure out who can be tied to wanting to take ownership.

Because that's the other part. Even if there is a physician that's the most common prescriber, they don't want to take ownership of this. So that would be really nice to kind of see.

So do you have a time line of when the next step of trying to develop the model of attribution?

MS. O'ROURKE: So that's something where our group are currently really trying to explore funding, and hopefully it'll launch in the near future, so it's good to hear that this is needed and hopefully useful to the field.

CO-CHAIR PINCUS: Connie?

MEMBER HORGAN: This is just a follow-up. This is sort of my question. Is this an ongoing committee, or is it one time?

MS. O'ROURKE: It was one time through

the last scope of work. We're hoping we can get additional funding to reconvene the committee and to continue to explore some of the questions they raised.

MEMBER HORGAN: Okay, all right, thank you.

CO-CHAIR PINCUS: I'm going to call on myself now.

Just two points. One is, this is a great opportunity to also disagree with Peter.

(Laughter.)

CO-CHAIR PINCUS: So yes, I'm a heretical heretic.

So anyway, while I agree that we should focus primarily on teams, but I think the key challenge is, who's on the team? Making the determination of who is on the team.

And part of it is also, once you've determined who's on the team, the question is, are they behaving or performing as if they're on the team?

So I think that it's not totally --

there is an individual responsibility to be part of a team and that they're playing with a team. And so that's part of the, I think, the accountability conundrum, is actually thinking -- figuring out who are the members of the team.

And. then it's whether is that an individual provider level or at an organizational level. So I think that's -- and so some part of your modeling should include that kind of consideration of how do you make that determination.

Secondly, I really do think that this should be part of the criteria that we go through. And that there should be some sort of, whether it's explicit attention to this as part of the validity component or is a separate component of the accountability model, should probably be integrated into what we've been discussing.

Because if you think about our discussions over the past day and a half, it's a lot of this comes down to, it's like, it's been

an accountability issue.

It's been framed under other categories, but that's really what it's been.

So then, Mike?

MEMBER TRANGLE: I think what I'm going to say is in the nature of comments.

I'm involved in at least three different types of activities where attribution is occurring. And so for one thing, we just, in our state, we had discussions about as more and more people get integrated, who gets the attribution for depression results, for example? You know?

And I would say it's becoming the norm in my area that at least the large systems of care are designing their outpatients' behavioral health systems, because of shortages and difficulty getting in, to limit it to you have to be an active patient of our primary care group to get in. So everybody we get is also being seen initially and probably treated a little bit by primary care.

And then we get them in our behavioral

health clinics, or we see them in the primary care clinics for co-locating and doing some kind of shared care, collaborative stuff.

And we've argued it back and forth. We want to sort of say it's okay to be -- to double count. And that in some sense, theoretically, if we're all part of the same team, baseball team or whatever sport you want to say, do we all rise or fall together?

But I know within my medical group, when you're looking at total cost of care, how much of that's coming from primary care, how much of it is coming from they got hospitalized with a heart attack, and so would that be attributed to the hospitalists? Or the cardiologists? Or both?

I'm involved in a number of discussions where it feels like -- in one of our regions, we've got a joint alliance with another large system of care, and we're looking at total cost of care. And we're doing a number of sort of joint things to sort of add therapists and some prescribers in our -- both our systems of care,

primary care clinics.

We've contracted with some outside groups who clearly are not playing on the same team, where their vision is not the triple aim, but their vision is more business.

But it's a large group and --

CO-CHAIR PINCUS: Or less business.

MEMBER TRANGLE: Well, but so like, they don't care about reducing readmissions or think twice about sending people to the ER. And we build in things to try and do that into our clinics.

And it just becomes sort of dizzying. And your point about, what are you talking about? What's the issue? You may have a different methodology or mix in the way of thinking about it.

Because it's just so -- there are so many different multifactors, what are the key ones? I get dizzy thinking about it.

But I also sort of get annoyed that it feels like you read in the literature about the

downstream effects and the indirect impacts of mental health treatment on saving money, but the way we measure our total cost of care, I don't -- we haven't been able to succeed at attributing it in a way where we're giving ourselves adequate credit for it.

It's all coming -- it's all sort of laying on the primary care docs who really don't know what we do or really care as long as we're there to treat their patient when they want it.

So it's much more complex and difficult when you're trying to do it separate from just theoretically talking about a measure.

CO-CHAIR PINCUS: Thanks.

So Charles, then Rhonda, and then Mike.

MEMBER GROSS: A couple of related questions. What's next with this?

The second is, the attribution model is not specific to behavioral health. And if it's not, then are there subsections related specifically to behavioral health?

And that goes to Michael's question a

little about, what about when there's shared risk or shared accountability for integrated care practices?

And then the third area is, is the model built with some thought to any potential differences based on the product mix that the practice may be addressing? Commercial, Medicare, Medicaid?

MS. O'ROURKE: Sure, so, as far as what's next, I think we're really hoping to explore ways to get the Selection Guide in greater use, to partner with some of our member organizations and others to hopefully get that out there to developers, implementers, even to clinicians and providers as you're negotiating contracts. What's a fair attribution model?

So I think anything -- if there's anyone interested in the report and getting that out there, we'd welcome some buy-in there.

We're also hoping that we could explore ways to better integrate an attribution review into our endorsement and selection work,

understanding that it's become a really key question that has come up both through our CDP and MAP processes.

I think we also recognize there's a need to better understand the pros and cons and the unintended consequences of what we can do to continue to see what evidence is out there about when a given model might be more or less appropriate.

We didn't get into exploring individual areas of care. I think the committee tried to keep it at a more general level, so there weren't necessarily specific considerations for behavioral health, other than recognizing in the report that it's an extremely challenging area for attribution, but no set answers there.

And then I think, finally, for the third question, the committee really grounded their recommendations in the need to understand the context of use and that a certain model might be more or less appropriate, depending on how it's going to be used, and what's okay for reporting

or internal improvement might not be fair for payment purposes.

So the report gets into some of the nuances there.

MEMBER GROSS: And just one of other comment. First, this is great work, and I support Harold with making this some factor in the NQF evaluation.

And then, just to close, I'm surprised you only found 160, because that must mean you only spoke to 160 people.

(Laughter.)

MEMBER GROSS: I'm just off the top of my head, even within Anthem, I can come up with half a dozen different variants of attribution based on specialty and provider type. And that's within government business, to say nothing of the commercial side.

So I think one of the complexities is going to be the inevitable friction when you look towards implementation, since all these are on paper, imbedded in contracts with providers, and

to change them is a significant lift.

And I know many companies and provider groups are already doing different versions of attribution models. So it's a great project but one that's going to be long-term, I think, but much needed.

MS. O'ROURKE: Absolutely. I think you hit on some of our key challenges for going forward. Our 163, I believe is what the authors could find that were publically available, and recognizing that there's even more that are in private contracts and not in the public domain.

So really reining in all of that is quite the challenge.

CO-CHAIR PINCUS: Rhonda?

MEMBER ROBINSON BEALE: I also want to applaud you taking on this work because I think it's very well needed.

Just as Charlie said, I think, need to look at it -- you talked about it from a product perspective, I'm going to talk about it from a different perspective, whether it's a PPO, okay,

versus other types of what I would say, structures.

Because the PPO, which is where a lot of our employer groups are still residing in that area and still want their members to have the choice to go where ever they want.

But yet, they want us to be able to measure the effectiveness of the system and of the providers so that the members can select the providers who produce the best outcome.

So in trying to do that, it becomes a very difficult situation because we're not to ask the member to choose a primary care. They may chose a whole system, or they may not. They may -- you may just look at their pattern of where they go and then attribute it silently to that provider.

And I have a lot of problems with that because I think it becomes real problematic in terms of people who like to go from one place to the next, and someone may have seen a primary care physician three times in the beginning of

the year and then went to see a specialist -- because there is no rules whether it's just a primary care or a specialist -- five times at the end of the year, and now they're being attributed for that year's outcome.

So there's a lot of problems with this that really need to be addressed. And I would really suggest that PPO is one of the areas where it's going to be more problematic.

Talked about behavioral health, so I may be sacrilegious, but I've already started this with behavioral health in my plan, where we have been introducing for the past year-and-a-half the whole issue of attribution but accountability.

Accountability is far more important. Attribution methodology is pretty simple. So for example, substance use disorders, we're saying that recognizing it's a chronic illness, recognizing that the longer one's in treatment, the better off they are.

We contract with I would call them maybe semi-comprehensive providers who can provide IOP

as well as outpatient. Then they become accountable for that population that they touch. If they see that person more than three times in their IOP, they're now accountable.

So any subsequent hospitalizations or anything like that is attributed to them.

We've also challenged our psychotherapists in the community with another NCQA accreditation measure, which has to do with people who have behavioral health illnesses having a wellness exam at least once a year.

So that didn't go over well when I first introduced that, but when we started explaining to the behavioral health providers, particularly the psychotherapists, that you see these people more than the primary care physician; you have more opportunity to convince them to get an annual physical. We, as the plan, have to get back information to them to let them know if that was completed.

So that may sacrilegious, but I've already started this experiment at this point.

CO-CHAIR PINCUS: Okay, Mike?

MEMBER LARDIERI: Yes, I guess I'm backing to just support what Mike was saying before, and I think, also on the teams, I think we need to talk about ritual teams.

And I think it gets real difficult and also crossing across systems. And so I think that's what makes it difficult because, in our situation in New York, we have these PPSs. It's mostly on the medical side. They pay downstream to behavioral health providers, but there's not a lot of that money coming downstream.

But everybody's working with that same patient, and it's across 10 or 15 different tax IDs that all these services that are happening. So it gets really difficult to manage.

And I guess it's all the other financial relationships that people have and how they structure it so they get a piece of that savings, that bonus payment, whatever. And that's like a real difficult thing to do so far.

CO-CHAIR PINCUS: Thank you.

So Erin, it seems like you've got a lot of support for the healthiness of the --

Oh, there's one more? Oh, yes, I didn't see yours. You're sideways.

MEMBER ZUN: No problem.

I think it's a wonderful concept. I'm just thinking from some of the experiences that I have as an emergency physician and patients coming to the emergency department as their primary care provider, how you would attribute their care.

Besides that, the sharing of or the collaboration between primary care and emergency care providers. How do we decide who's responsible for what?

It's a very complicated model. I guess you guys are trying to make it really simple. I'm just not sure how to apply it across the care continuum.

MS. O'ROURKE: No, that's a -- it's a great point. I think that is probably one of the committee's chief findings is this is not simple

and will probably never be simple.

And really, they kept coming back to the idea that each situation needs to be individually assessed and what model might be most appropriate could change.

But there was a recognition that dealing with say 163 different models is causing a lot of problems. So if there's a way to standardize and allow for best practices to emerge, it could hopefully reduce some of the noise out there.

CO-CHAIR PINCUS: And I think this accountability for the emergency room providers is among the most difficult conceptually when you think about it.

So well, thank you. This has been very helpful and we look forward to further discussions. Because I think this is an area where there may be -- it may be worthwhile to actually do a deeper dive on sort of the accountability models within the behavioral sphere.

MS. O'ROURKE: Absolutely. Thank you

for the opportunity and your great feedback. This was really helpful, and we'll -- I'd be happy to keep the committee in the loop if we move forward with additional phases of this work.

CO-CHAIR PINCUS: So Tracy, do you want to lead us through a kind of look across the span of, I guess our committee's responsibilities with regard to behavior health measurement, and thinking about both gaps as well as how we sort of harmonize across the different measures and categories of measures?

DR. LUSTIG: So I think we talked about this very briefly on our orientation call. But one of the things I had tried to do for you here was try to take a step back at a few thousand foot level and look at our portfolio overall.

I hope I don't confuse you more than enlighten you. But what I tried to do was look at our entire portfolio of measures and put them into different categories and try to give us a perspective overall of what our portfolio looks like, either by the diagnoses that we're focusing

on or the part of the care continuum that we're focusing on, or those sorts of things, just to hopefully stimulate you think about where you think the gaps are and where maybe measure developers should be focusing.

So this is kind of our standard opening slide about our portfolio, but obviously we're charged with looking at behavioral health conditions, measures that can be used for accountability in public reporting for all populations and in all settings of care.

Some of the common topic areas that our measures to look at are alcohol and substance use, tobacco use, ADHD, depression, schizophrenia. And we currently have more than 50 endorsed measures within the area of behavioral health. I think this number still stands about right after our meeting today.

When I did look at the portfolio overall, I was considering all of the measures that we have looked at over the last two days.

I'll also say, as you can imagine, you

could fit in any of our measures into more than one category, even when just looking at diagnosis.

So this is sort of just like a rough overview for you.

So the first thing I tried to do was look at the main diagnosis that was being paid attention to in each measure.

So how it kind of broke out when I did this roughly was physical health, which, what I mean by that is, we have a whole host of measures that are really around things like: if you have schizophrenia and you have diabetes, have you had your neuropathy exam?

So I kind of categorized all of those with really the focus was more about ensuring that you're getting the physical health needs assessed if you're a high risk.

Depression focus came out at 19 percent; tobacco use 15 percent; alcohol and other drug use 15 percent.

CO-CHAIR PINCUS: Is each measure

counted once in those?

DR. LUSTIG: Each measure is counted once. So I made a decision of which one it fell into more than others. So yes, there's no duplication.

But like I said, there's a little bit of wiggle room, because there were cases where I sat there and thought, given this --

That's why you'll see these other two categories also. 10 percent of specific diagnosis, this would be things like the measure we looked at yesterday about ADHD, where it was a very specific diagnosis, but there were really no other measures in our portfolio that were that exact same specific diagnosis.

And then when I put general behavioral health, this is where often it was things like serious mental illness or other things where there were more than one, or it was things like experience of care where it would apply to kind of all the diagnoses in our portfolio.

So this was one way that I looked at it

to try to help you think about it. The next way I tried to look at it was, well, within some of these larger diagnoses, what can I pull out to try to further break down?

So you'll still see things like physical health, depression, tobacco, AOD there. But then the ones that I really tried to pull out here were medication use, 20 percent of our portfolio.

And these were things like we've been talking about today, certainly: adherence to care, appropriateness of the choice of medication, continuity of care, but specifically focused on medication use for whatever the diagnosis was.

The other thing I pulled out was care coordination, about 9 percent of ours. These had to do with things like follow up after a hospitalization.

So I was trying to think about two different settings that would have to coordinate for the measure. And then our few experience of care measures that I pulled out there.

And I'll just kind of go through these quickly. And again, I was just hoping that it would stimulate you to think about what's missing here, or what are we concentrating on too much or not enough?

And so then the last way I looked at it, you may or may not remember this from our orientation call, but I had also tried to think about our portfolio in terms of where in the care spectrum, as I would call it, are we -- are our measures really focused?

You can think about the population at risk, so that would be a lot of screening issues. Evaluation and initial management, I tried to think of that as, you know, what are you doing at the initial diagnosis? And then the follow-up care has to do with continuation of care and continuing to follow up.

So when I --

(Off microphone comment.)

DR. LUSTIG: This borrows from another group that had used this, but it was meant to

show that along these, depending on the care, people can go in different trajectories. But sorry, I should probably know that.

Yes, Raquel?

MEMBER MAZON JEFFERS: Did you also distinguish between process and outcome measures?

DR. LUSTIG: That I did not do for this. I'm sorry. But, I can definitely follow up with that. I can certainly tell you that a majority are process measures that we have like with every portfolio, I think.

So this is --

CO-CHAIR PINCUS: So it would be useful to actually have a discussion about -- at some point, about what is the state of the art for outcome measurement, in terms of -- and to what extent do the existing outcome measures demonstrate that state of the art?

DR. LUSTIG: And, so, when I was looking over this, this is my very crude -- my art skills are not as good as they should be -- but, this was my attempt to show where, again, in that

breakdown.

And what this really means is that about 22 percent of our measures are focused on that screening or that, you know, looking at the population at risk.

I have the overlap there because we have a certain number of measures that are about screening and if you find something, you initiate treatment. So, that's where those roughly 16 percent of measures fall.

And, again, nothing is counted twice here. So, the ones -- yes.

Can you use the microphone? Sorry.

MEMBER TRANGLE: Right now, you want it on? But I did see it.

I think we're getting sort of the beginnings of what's going to be a growing category of prevention, you know, versus just screening.

If you think about some other point of first episode of psychosis, you know, or some of the health measures around people that have SMI,

how do we prevent the metabolic syndrome and what's going to happen down the line from that?

Whether it's they don't become so obese so you can prevent them from getting diabetic, you know, things like that.

DR. LUSTIG: And just to finish it off, it really was -- I categorized about 20 percent of the measures falling in that. You've received the diagnosis already, so how do you initiate the treatment?

And then, that last group, the 40 percent, 41 percent, the measures are really focusing mostly on follow up. So, you've already had a diagnosis, treatment has begun and what are you doing to follow up with it?

And then, I tried to put across the bottom here just that experience of care that's crossing all of those.

Yes, Mady?

MEMBER CHALK: We've got to get more clarity and subdivide that follow up category. Follow up can mean many things.

DR. LUSTIG: Absolutely.

MEMBER CHALK: It can mean follow up with the primary care physician. It can mean follow up in treatment. It can mean on and on and on. So --

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

DR. LUSTIG: And we can do that --

CO-CHAIR PINCUS: Yes, because it's -- that's the most surprising thing because I don't think about of having a lot of follow up care measures.

DR. LUSTIG: Sure.

MEMBER CHALK: But I think sometimes we don't know what we're talking about.

DR. LUSTIG: Yes.

CO-CHAIR PINCUS: Yes, like, you know, seven-day follow up after a hospitalization, it's not --

DR. LUSTIG: Right. And, again, this is not a perfect science at looking at this either. This really was meant to give you more of just a

general overview of where our measures are falling.

I fully agree, I sat there arguing with myself sometimes about which place a measure would go. But, at least I thought this would give us a rough idea of where our part fully sits.

CO-CHAIR PINCUS: So, Les and Connie and Peter?

MEMBER ZUN: Yes, before I have to excuse myself.

So, I'm kind of wondering whether the tail's wagging the dog or the dog's wagging the tail.

Because, you know, it seems like NQF is dependent on all these other organizations to bring measures to us rather than saying, here's the measure we think is needed.

And maybe I'm mistaken, but I think we discussed some issues concerning integrated healthcare and those kind of things.

And from my perspective as an emergency

doc is, to me, all the patients that come to the ED because their care network has failed them.

And we always look at the measure of, you know, their insurance status and that's the way to capture all the data. And we've got to make it easy to capture data.

But every day I take care of patients in the ED because they can't get their meds. They can't get follow up. They're in a care plan that just let them fall through the cracks.

So, I'm not sure we're getting at the, to me, some of the measures that really impact on patients' lives that have chronic mental illness.

And I really think that we kind of need to turn this on its head and say, what do we think are needed, what are we, as a group, think are needed and why don't we go to all those organizations and say, we think you need a measure to determine whether the -- why the patients are coming to the emergency department? Or, why they can't get their meds? Or, why they can't get therapy? Or, why they can't integrate

their medical and psychiatric care?

Thank you so much.

DR. LUSTIG: I can say we agree and NQF is definitely making steps -- taking steps to get more into the driving toward the measures that matter, as we call them.

I think some of you have heard about, and probably something we should have maybe had a presentation on is the Measure Incubator which some of you may have heard about, but is really our effort.

We can't develop measures, but we can act as a convener to bring people together who have ideas of where the measures should be and start to help them think through those processes.

And do you want to mention about that our strategic plan --

MEMBER ZUN: And I'm happy to volunteer and so are the organizations that I represent.

MS. MUNTHALI: Great, so we'll follow up with you because we are facilitating the development of measures. It is part of our

strategic plan, not just by having the folks that have the ideas, the concepts, but those that have the technical knowhow, so the measure developers and the data. And those that know how to test data.

So, again, we're not developing measures, but we are convening the folks that do this sort of work.

The great thing about it is, we are getting to those measures that matter through the incubator.

A lot of the first measures to come through are patient reported outcome performance measures around, we know that there are gap areas.

What we're trying to do is to spur measure development in innovative areas where there's significant gaps, behavioral health is one.

If you go and look up our incubator work, you'll see that behavior health is listed there. That's deliberate because we know that we

have significant gaps in that area.

And so, I think we can share post-this meeting, information about that. But, you are very right. In the past, NQF has just kind of waited for measures to come to us.

Peter can tell you a little bit more about our strategic plan, being on our Board as well, where we are beginning to focus on those measures that matter and the gaps that matter.

And part of that is to have our committees, like your committee, to, you guys know where the gaps are. You know where we need to focus, to help us to think about, you know, what are those leading indicators of health within the context of behavioral health, what are those drivers that can get us to those indicators.

So, this is just an early exercise for us to start making decisions about how we are saying that these are the priorities in behavioral health.

What are the measures, quite frankly,

that don't move us towards improvement and are increasing the burden with regards to reporting and data collection?

So, we're beginning to empower ourselves, beginning to empower the committees, so more to come.

Peter, I'm not sure if you wanted to mention anything.

CO-CHAIR BRISS: I'll --

CO-CHAIR PINCUS: Yes, Peter was next and then we were going to go down this line and come back up.

CO-CHAIR BRISS: All right, so, I was going to ask for another slide that I'd like to see eventually is the -- it sort of reflects an issue that we've been talking about a lot this meeting with behavioral health integration.

So, I'd actually like to see what kinds of providers and settings are affected by our measure portfolio. My guess is that it's heavily -- it's still heavily skewed toward kind of specialized behavioral health settings.

And if that's true, I think there's a case to be made for more coverage of primary care. I'd be happy to be wrong about that, but I suspect that there's a whole lot of behavioral health stuff that happens in primary care settings that if we don't have a lot of measurement, we ought to have more.

CO-CHAIR PINCUS: Although, actually, I think we'd find that it's actually fairly balanced, but it's not -- but, if you then stratified it by the quality of the measure, it might not be.

So, Mady and let's work our way down here and then come back up the other side.

MEMBER CHALK: Now, I don't think you and I agree about that, Harold. Yes, there are a lot of physical health integration, serious mental illness measures, but no discussion and no measures here that have to do with substance use, disorders and primary care.

Despite the fact that there are all kinds of -- on these measures that are in our

thing here, this is all serious mental illness and physical health.

CO-CHAIR PINCUS: I'm not sure what you're looking at.

MEMBER CHALK: Within our folder, it is behavioral health portfolio.

DR. LUSTIG: This is in your, you know, an extremely shortened title, but meant to give you a very --

MEMBER CHALK: No, I'm just looking at this physical health --

DR. LUSTIG: Yes.

MEMBER CHALK: -- all this list of things, matches, that's all. I'm assuming everybody had the same materials.

And all I'm suggesting is that screening for physical health, if we're going to -- I want some discussion of if these are measures for schizophrenia or serious mental illness, why are they not for alcohol or drugs and what are we -- we're going to have 400 --

Now, I should have a developer show up

and say, okay, for all of these that are listed here for schizophrenia, it should say alcohol and then drugs and then we'll throw them on a provider or a health plan and say, now you've got 42 different measures to respond to instead of 24 when you could have one.

CO-CHAIR PINCUS: What do you mean by that?

MEMBER CHALK: What I mean by one is screen for physical health for any major mental illness or list them, schizophrenia, bipolar --

CO-CHAIR PINCUS: You're talking about a composite measure.

MEMBER CHALK: Okay. However you want. This says behavioral health up here, it isn't behavioral health, it's mental illness. It isn't behavioral health.

And I'm saying why doesn't it say alcohol, drugs, however you want, a diagnosis?

CO-CHAIR PINCUS: So, let me just interject something that -- so, we've had a number of discussions about thinking of people

with fill in the blank, in most cases, serious mental illness --

MEMBER CHALK: Right, okay.

CO-CHAIR PINCUS: -- as a disparities category.

MEMBER CHALK: Okay.

CO-CHAIR PINCUS: And that being a potential strategy for sort of potentially low hanging fruit for measurement of sort of at this sort of integration interface with physical health and behavioral health.

MEMBER CHALK: Right.

CO-CHAIR PINCUS: And it -- because, you know, because the measure's already being collected on a general health basis and it's simply a matter of stratification based upon diagnosis which is readily available.

And there are two issues, though, that come up once you think about that.

Number one is, does that constitute like a doubling -- if you add subsidies, does it become a doubling or a quadrupling of every

measure or is it simply that you report measures in a stratified way? So that's --

But, I think this --

MEMBER CHALK: Okay, that's fair.

CO-CHAIR PINCUS: Okay, but I'm not arguing with you.

MEMBER CHALK: That's a fair question.

CO-CHAIR PINCUS: I'm just --

MEMBER CHALK: Yes.

CO-CHAIR PINCUS: -- trying to summarize.

And so, that's something that NQF should think about is, if that's the case, does there need to be a whole process and a separate measurement category or can it be something that's kind of routinely reported, you know, on certain designated measures?

MEMBER CHALK: Good point.

CO-CHAIR PINCUS: As you might do, you know, look at, you know, sociodemographic status or something, not as a risk adjustment tool, but as a way of actually stratifying to see

performance.

MEMBER CHALK: Good point.

CO-CHAIR PINCUS: So, that's -- and then, if so, then that raises two further questions. One is, how do you define that segment?

MEMBER CHALK: Okay.

CO-CHAIR PINCUS: And number two is, which of the preventative health measures are most appropriate to put in there, you know, based up the likelihood that this is going to be an important issue relative to that population?

MEMBER CHALK: Right.

CO-CHAIR PINCUS: Or, for that matter, if there's -- if you're doing it on the basis of chronic disease comorbidity, which chronic diseases, should that be like sort diabetes ventures?

MEMBER CHALK: Right, diabetes and hypertension.

CO-CHAIR PINCUS: Yes, so that's, you know, I think that's a fair comment. But, to

actually have a systematic process for doing that.

MEMBER CHALK: But we need that.

CO-CHAIR PINCUS: Yes.

MEMBER CHALK: We need that.

CO-CHAIR PINCUS: And it may not be something that requires a whole measure development process because the data exist there, the measure's already being collected and it's something that could be done, you know, relatively easily, you know, with an additional analysis of the data.

So, that's something to think about. You know, almost having a different track for that, that would help to round that.

And then doing some testing, though, particularly if you want to make a composite which might make an easier thing to digest and use.

So, I think, you know, so, just sort of summarizing what you were thinking about, that might be sort of a recommendation as we move

forward to think about how we might do that.

They might, you know, in some ways nest more of these measures under a single heading of a measure without creating new measures.

MEMBER CHALK: That's right.

CO-CHAIR PINCUS: Yes.

MEMBER CHALK: That's right. That's where I'm heading, I think.

CO-CHAIR PINCUS: So, Raquel?

MEMBER MAZON JEFFERS: So, Tracy, I don't know if you had more slides, but it would be helpful for me to think about gaps if I could also see other cuts of the analysis like process outcome which I said before, care settings which was mentioned.

Also, is the measure patient report, is it provider report? Is the measure at the level of the health system or at the level of the provider? You know, at what level -- at what --

Because I think we have a lot.

CO-CHAIR PINCUS: Shall we dare say accountability?

MEMBER MAZON JEFFERS: Right, right, that's right.

I think we have a lot of measures that are at the health system level, but I'm not sure how each of them fall out. So, that would be helpful.

And then, I wanted to make a comment picking up on something that Mike had raised earlier and we agreed to defer to this part of the day to talk about, and that was, particularly in acknowledgment that much of the opiate overdose epidemic has been driven by primary care physician over prescribing of opiate medications for the treatment of pain.

To what extent is it this committee's responsibility to also look at prescribing -- opiate prescribing --

CO-CHAIR PINCUS: There is a measure that was endorsed last time around.

MEMBER MAZON JEFFERS: I don't remember that.

CO-CHAIR PINCUS: I thought there was

one.

MEMBER TRANGLE: Get a handle on indications and then overusing, how do you start triggering and capturing overusing their patterns like that? You know? So, it's like monitoring and then catching and hopefully preventing things from getting too out of whack or catching them earlier versus you've already got the new one is find an easily discernable way we can measure use of a med, you know, after the fact?

MEMBER MAZON JEFFERS: There are now from the CDC nine evidence-based practices for the -- for better pain care. And they would not be hard and I know there are some groups already working on trying to construct measures that align with those best practices around pain care.

And I realize we're really talking about primary care providers prescribing practices for pain which is I guess not a behavioral health issue, but it's so -- but it is very connected to the opiate overdose issue.

CO-CHAIR PINCUS: You know, just a

comment. When we've looked at it, it's more than primary care, dentists give out stuff.

MEMBER MAZON JEFFERS: Oh, dentists?

CO-CHAIR PINCUS: It's a post-surgery.

MEMBER MAZON JEFFERS: Dentists?

CO-CHAIR PINCUS: I mean, so there's like -- there are, in fact, several measures looking at opioid use, opioid prescribing, yes.

Okay, so, they're not in our portfolio apparently.

MEMBER MAZON JEFFERS: Right, a mistake.

CO-CHAIR PINCUS: Yes, so, I think -- so this also, you know, raises the issues about how we're defining the net -- for casting the net of what's a behavioral health measure.

CO-CHAIR BRISS: Yes, and it -- remember that there's never going to be a clean -- people don't fit, real people don't fit into neat, single boxes. And so, there's never going to be a clean, unarguable sort of definition of what's my turf.

NQF is already doing some -- trying to

do better than it has done about having committees talk to each other on relevant things, you know.

And so, one of our -- one of the tobacco measures that we were talking about this couple of days actually came out of the whatever we call pediatric or child health or it's whatever.

And so, if it's childhood smoking, no reasonable person could call it behavioral health. A reasonable person could call it child health. It's -- there's not going to be one answer about where it goes. The committees just need to kind of coordinate what they're doing.

MEMBER CHALK: Yes, and it would be good if we could, particularly on the pain issue, on the opioid prescribing issue. If we could at least be informed that they're talking about it.

CO-CHAIR PINCUS: Yes, I remember because I came through the MAP Committee. And, so, I'd assumed it was there.

But, anyway, Tami?

MEMBER MARK: Yes, I just want to

emphasize, you know, I think it's kind of like you said, low hanging fruit to extend some of these measures to substance use disorder, post discharge follow up, some of the screening measures.

But if they're not extended, people are reluctant to use them. And I found that when I talked to CMS and states. I said, well, this is basically, you could just use this measure to substance abuse disorder. Oh, it's not endorsed.

You know, so I think there is something you said for putting it in your portfolio and finding some easy way to do that.

CO-CHAIR PINCUS: Mike?

MEMBER LARDIERI: Yes, and I'm going down the same road as Mady. I think that if we can get the developers to work together and then, you know, stratify the responses. Like, it's one measure, why should you have 20 different cuts, you know?

Maybe that gets into, you know, what I was talking about earlier, the economics of

bringing those developers into the incubator program. And, you know, next time you're not going to have ten or five, you're going to have one.

You'll have to work together and figure it out, which is a difficult thing to do but if you don't do that, now we're sitting with 50, 60 when out of the 50, we could probably have 30 because a lot of those could be collapsed into one if you just stratified some of these responses.

In the opiate discussion we had this morning, they kept saying, well, the initiation, that's another measure. Well, it's not really another measure if you use the initiation and then the use of the opiates. And then ask the next question down the road, you know, who's opiate free at a certain period of time?

Now, we have an outcome measure and you have all the steps you need to do to take it, maybe we need to think about these are all the steps you need to take in order to get to the

outcome. Make it one package as opposed to pieces all over the place and that's sort of how it is right now.

So, I would go for being stratified and bring those folks together. I know it's an issue with how it works, but I think we need to get past that otherwise, it's hard.

CO-CHAIR PINCUS: Dodi?

MEMBER KELLEHER: So, I think my comment is to even broaden out more than the individual issues that are coming down the table that I think is in sync with it.

One of the things that became a theme for me over the last two days, not only with, but especially with the maintenance measures is the amount of time that goes by and they come back with a measure that not only is sort of on autopilot for the face validity, but has not kept up with the practice, whether it's the clinical practice or the business practice in the last three to how many years.

And that we're looking at if there's

still a gap, so a reason to continue it. But, that in and of itself seems to be insufficient anymore.

What we need to be also -- and the fix may be to advise ahead of time which is more work for NQF, you know, what they need to look at beyond whether there's still a gap.

So, and again, this is not the only issue. For instance, there was a measure where, you know, they're excluding behavioral health done in a PCP's office when we know that 80 some percent, maybe even 90 now, medication for an antidepressant is done in the PCP's office in the commercial practice.

So, it makes -- it doesn't really reflect the actual practice as it exists today.

You know, I could pick out a few more, but, you know, I think people will probably know what -- I hope you know what I'm talking about.

And so, somehow I think that we need to address that.

CO-CHAIR PINCUS: So, you're saying

essentially to place a higher bar for maintenance measures?

MEMBER KELLEHER: Yes, that they can't just come back without having, you know, also, you know, reviewed, you know, their exclusions and, you know, how they're going about it in a way that is more in sync with what the current clinical and business practices of the people are going to have to deal with all these measures once they're introduced.

CO-CHAIR PINCUS: You know, I would agree with that and just to -- that's come up in the MAP meetings where -- Measures Application Partnership.

So, just be aware, NQF is basically two lines of business. One is the standing committees for consensus development, but the standing committees for, you know, because that's development like this one.

But then, there's a component of the Affordable Care Act that requires that there be a nongovernmental, multi-stakeholder group that

reviews all the quality measures used by CMS.

And where there's a recommendation about whether or not those should be used in the programs of the ACA and other related programs. And so, that's -- so CMS every year proposes it.

So, anyway, that -- my point is that in the MAP process, there's been a greater request to get feedback from CMS about how the measures have been used, what's been the performance? What have been the problems encountered? You know, in a more formalized way.

And I think that would apply here as well that, rather than just update the same form that actually have a somewhat different form for a maintenance measure that would actually include, you know, who's been using it? What have they found? What have been the problems?

So, I think that that's, you know, I think, Dodi, that's a very useful suggestion.

CO-CHAIR BRISS: And fair warning, as Harold and I can tell you, you're eventually, if you're around these tables enough, everybody --

you get sucked into all of the available lines of business sooner or later.

CO-CHAIR PINCUS: Yes, yes.

MEMBER KELLEHER: So, my only other comment is really not about the portfolio per se, but I want to make sure I get it in, so I don't like to have to talk too much.

I would find it useful, and I think the committee would as well, it might make some of the questioning and the decisionmaking a little less torturous on some of these cliff hanger ones.

If we had a little bit more pre-orientation or maybe for those who are going to be on the next phase, maybe not right before, maybe it's, again, it's more work for you, it's over the year or two before everybody gets that together, around some of the stuff that you may assume we all have a sense of but don't.

So, we were talking about how you can endorse a measure and yet have specific, you know, essentially saying don't come back unless

you deal with this on a maintenance measure, but, you know, strong statements of required.

And then we got into the discussion about whether you put an asterisk there or not.

That's the sort of thing that may have made it easier, would have made it easier for me if I'd known that before, right? So, I see a measure, I think there's still evidence for its need. I think it's not perfect, but it's good enough, but it really -- I don't want to see it again unless it addresses X, Y and Z.

And so, that's -- there's the sort of orientation that, rather than you get it if you sit through one of the phases, but maybe it'd be easier to sort of anticipate and the orientation be a little broader than that.

That's all, I'm done.

CO-CHAIR PINCUS: Andrew.

MEMBER SPERLING: Thank you.

So, this is my first meeting and it's been very enlightening. I'm the rookie here.

I've learned a lot and I'm hoping to be

able to contribute a small amount. And I now know that I have a lot more to learn about the measure development processes in NQF.

The fact that, it is a -- and protractor is probably an unfair word to use, but it belies on, you know, validating at every point the scientific evidence backing up a measure, its use, its utilization, the gaps.

And I recognize there's a lot of research that has to be go into developing that.

My sort of ambition in getting -- becoming a part of this panel was to try and push NQF, particularly relative to measures on psychotic disorders, schizophrenia, schizo-effective disorder, bipolar disorder, particularly around quality measures.

And not that the process measures that have been developed already aren't important. We know about the, you know, the 20 years of early mortality experienced by these patients because of poorly managed diabetes, asthma, heart disease, on and on and on.

We know that's a public health crisis and we know it needs to be addressed.

But we really, from NAMI's perspective, we really want to see this body move toward the development of real quality and outcome measures, particularly around schizophrenia where we believe there's an enormous void out there.

So, I look forward to continuing to participate in this and figure out a way to work with you, Howard and Peter to figure out, to work with the staff and begin that, you know, plant that seed, get that initiation of that process going around development of quality and outcome measures, particularly around psychotic disorders.

And I look forward to continuing to participate. And, so, thank you for allowing me to be here and I look forward to the appropriate people on the NQF staff, if you can recommend and Tracy, and I'll follow up with that for this meeting to sort of begin the process of planting that seed.

Thank you.

CO-CHAIR PINCUS: Shane?

MEMBER COLEMAN: I'll try not to pontificate too much or let myself get too out of control, but, I'm going to come from a little bit of an idealistic perspective. And I'm not sure this totally falls within the scope of NQF, but I guess I just -- I'm going to say it out loud anyway.

So, I think that -- oh, let me also say that I come squarely from behavioral health. So, if any of my comments sound like I'm down putting behavioral health, I'm including myself in that group. So, don't take it that way.

So, I think that something that I'd be really interested in is backing up even a little bit more and really creating a roadmap of sorts for behavioral health.

I mean, with behavioral health merging into medicine, the difference in cultures and the difference in the starting place of where folks are with regard to measurement of health and

outcomes, things like that, behavioral health is at a very different starting point right now than medicine is.

I mean, it's complete -- it's a chasm, okay, between the two places. If you, on the ground level, go into a behavioral health clinic in rural wherever, they don't -- and again, I'm totally putting myself on the spot there, okay, and I'm speaking generically and several types, blah, blah, blah -- but, the starting point is just very, very different.

And so, what would be very useful is an actual roadmap, literally, to what does it mean to measure behavioral health? What should the focus areas be?

I mean, and I think about, you know, just to be kind of concrete, things like functioning, you know, assure of symptom measurement.

And a lot of what we're talking about fits into what I'm describing. But, I guess, to just even endorse a bigger picture of like, hey,

these are the areas you should be thinking about if you're going to start thinking about measuring behavioral health in your healthcare system.

Here are some, you know, key measurements that fit in each of these areas. But, really a roadmap to begin with to say, oh, you know what, we should be thinking about measurements in these five core areas.

Oh look, they have, you know, ten measurements in each of those core areas. They have these ones starred because that's the plus endorsement, they think these are the best ones in those areas.

Like, that kind of roadmap would really help folks really think about quality of care that they're providing and be able to, I think, deliver it in a more meaningful way.

And lastly, I won't go into why I have this perspective. I mean, I feel like I have a fair amount of experience with behavioral health folks starting to grapple with this even I've found this is CCO stuff in Oregon, but we deal

with it a lot back home, et cetera. So, I'm trying to incorporate some experience that I have in this area.

But, I think finally, I'll just say that, even coming up with a roadmap like this, if you, you know, like the fact that you've looked at this in three different ways, I think three different slides, right, you started with diagnoses and then you went to kind of the continuum of care and then you --

Even being able to say like thinking through these slides and saying, no, no, these are the key pieces and then in these key areas of sorts and at the key levels.

Like, literally, in my organization, we have division level data, program level data, provider level data. Right? I mean, but it's same exact process that everyone's thinking about in this room.

So, if we could put kind of all of that into a roadmap of sorts that endorsed, that's meaningful, it would just, I don't know, it would

be an amazing resource nationally for behavioral health. I think it would be a big deal.

CO-CHAIR PINCUS: Mike?

MEMBER TRANGLE: I tend to agree with you, but that's not why I raised my name tag.

I have three year -- I have divided up four things I want to say.

I think there's a category of things that we should be thinking of doing and maybe the incubator is a place to do it.

Just in this last day and a half, we have two examples. You know, we talked about a measure where we're screening for depression. You know?

We also have a measure about we're treating and what percentage of patients are responding or reaching remission?

I think the absence in the middle there is somehow we should have a measure, once you've screened somebody, what percent and how do you increase the percent to get into treatment? You know? And sort of complete that gap in the flow.

You know?

The same thing exists for tobacco that we've talked about. You know? We've screened for tobacco. We're going to see if we give them a med, but somewhere, we want to measure whether or not they stopped smoking and what's the duration? You know?

So, when we've got these dispirit measures that could be important measures like both of these, we should fill in the gap. And I think the incubator should take that on to help us think about, it's a key thing to do so they really flow and become more meaningful. You know?

And for some of them I might be willing to volunteer. So, that's a category.

And I'll just give you the two examples that came up just now, but I am sure there are others.

Another -- I want to talk a little bit about two other areas that I think are not gaps, but like the Grand Canyon. You know?

One is, there are more mentally ill people in jails than there are in all the psych beds, psychiatric beds. You know? I our area, we're a fledgling, the counties that I would with and stuff doing stuff.

And some of the things we're looking at is are people being screened for mental illnesses? Do they get any modicum of treatment? And do they get connected with resources when they leave? You know?

I think we should somehow take on that as a huge zone that's playing out nationally everywhere. You know? And not ignore it.

Whether that's incubator material or something else, I don't know.

The other area, the other thing that's not the Grand Canyon, but maybe it's something a little smaller, but most places, and it's certainly true in our area.

So like, when we look at our state, I can't remember the exact number, David, maybe you would, but if you look at people that are

homeless, and every year we measure that. If you look at the ones that have mental illnesses and chemical dependency, it's risen every year.

It's up to either 62 or 64 percent, you know, of people that are homeless or people that have our two categories of stuff. The other large category is chronic medical illnesses, too.

But I don't see that being addressed. I see sort of, kind of the housing or access to housing or doing something with that as being -- it's not even in the purview of most of our Department of Human Services in our states.

It's in a housing bureau where our people that are sort of couch hopping and living beneath bridges and stuff don't even meet the criteria. They become homeless and get funded for stuff unless it's temporary bridge funding. But, it's a huge issue for our patients, I think, housing.

And then, the last thing is something that we've been grappling with a lot and are actually are measuring it and doing some

statewide studies.

But it has to do with, traditionally, the only places that measure access to, I think, access to mental health services are health plans where they're mandated to do it.

And if you're commercial in some places, you know, if there's a real shortage, the health plan will pay places to get people in in a timely manner, at least psychiatrists and prescribers.

But I think the lack of access which leads to sort of the lack of flow of psych patients is a national problem everywhere, you know?

And I think we should be taking that on and looking at that. You know? In our area, it's sort of, we've been, actually, for a number of years, we're measuring sort of how long people sit in ERs. You know?

What percent get admitted? What percent don't? But, you know, and we're also measuring the people that get admitted which, in our area, it's a little less than half.

But then, statewide, 15 percent of our psych beds are filled with people waiting -- who are safe, don't need a hospital level of care, but need an intermediate level of care, group home, a foster home or a crisis bed, something like that. So, that's 15 percent of our beds statewide.

And then, we're beginning to try to look at we need to increase our group homes and things they flow to.

But the flow of those psych patients, essentially, is ending up with them sitting in ERs and ending up in jails. And I think it's a huge national issue. And if we wanted to be aspirational, I think we should be thinking about that in our incubator.

CO-CHAIR PINCUS: Connie?

MEMBER HORGAN: Maybe this less aspirational, but I would like to extend two points that were made earlier.

I'd first like to extend what Peter spoke about of the emphasis on behavioral health

integration and let us not look at integration between mental health and substance abuse. I think we're overlooking that and I think it's really a key issue given the amount of comorbidity that's around.

The second extension that I would like to make is a follow on to both Mady's and Harold's comments related to, actually to Tracy's presentation on follow up and we're looking at follow up as the flow through the system.

And the examples that were given were more front end that when they follow or in the process.

But I'd like to link that back to what we spoke about this morning and the importance of including recovery as part of that follow up and making sure that that's a priority for us as we move forward.

So, that's it.

CO-CHAIR PINCUS: Mady and then -- oh, Peter?

MEMBER CHALK: Okay, I just want to

extend some of the things that Michael said.

So, we do have a screening and engagement measure for substance use disorders which amounts to screening and then did you get in to treatment somehow or other.

What we don't have is trends. We don't trend that measure. I mean, we say, you know, the comments, it's sort of a snide comment that people now, well, the measure hasn't moved.

Well, there are a whole lot of reasons why the measures don't move and we could have that discussion another time. I don't want to get into it.

But screening and engagement in treatment could be another one of those across the board kind of measures and trended.

With some, I don't know where that responsibility would lie inside the NQF, but certainly, I would like to see if measures are coming up for maintenance, what's the trend? And what have you done with it?

If you screen for depression and then

somebody got into treat, okay. But is that one percent of people? Five percent? Ten percent? And did it get better? Did it increase?

With regard to housing, CMS -- I'm working on a lot of projects with CMS right now and housing has become an issue as part of behavioral health and as part of the portfolio that has to do with patients with complex need and is more and more being included.

As a matter of fact, there are portions of the Medicaid benefit now that can be used for housing. It's a little crazy to try to get into the details of that but it's there.

And in the Duals Committee, which I've been on for a number of years, that pushed heavily the other services that are needed beyond health. And I do think this committee should get into it.

The other thing that's a big deal that you mentioned is this access issue. And I don't know how we would deal with that, but for CMS and for the work that I'm doing with them, provider

capacity has become a very, very big issue.

And whether -- how you measure provider capacity, what do you really mean by that? And quality of providers. But a provider is not a provider is not a provider.

And I'll leave it at that.

CO-CHAIR PINCUS: So, Peter, Raquel, Tami, Mike and then I'll sum up. And then, we've got to move on to an overview of the tobacco stuff.

CO-CHAIR BRISS: On the housing front, just for information, there are some places that are really -- yes, sorry.

On the housing front, there are actually some places that are now doing really innovative things, Durham, North Carolina, actually looked at their frequent fliers in emergency departments and across hospital systems and found that they had a definable population of people that kept showing up at the ER and they could save money by kicking in for some housing intervention as opposed to spending a lot of money on

uncompensated care in emergency rooms.

So, there are some models that are actually trying to deal with that that could be generalized.

CO-CHAIR PINCUS: Well, it's actually -- today's New York Times has an article about Canadian efforts around housing first in Alberta.

CO-CHAIR BRISS: So, but the point that I really wanted to make here were I'm delighted that we're looking at more of the portfolio as a whole.

And I think that NQF, this committee and other committees need to do more of that. I think that we need more attention to kind of families of measures and perhaps less attention to one measure because this is probably an extension of what Michael was talking about, and it's an extension of the last RAND measure.

Today, part of the problem with that measure was it's probably a relatively low-value area of an important health problem where we have better options where measures aren't available.

And so, part of the discomfort, I think, was that we were kind of addressing the wrong problem. And we don't have a good way to deal with that. We ought to do better, I think.

There's been a lot of conversation around this table that I've been really pleased with about the general burden thing. And, so, when we get to this stage of a meeting, we always come up with our top three or five additional things that we'd like to do.

And I'd like us, every time we do that, to come up with a matched set of three or five relatively low value things, at this point, that we could get rid of.

And because, part of the problem with NQF today is we've got too many measures and probably too many measures that aren't likely to really move the needle on population health.

And one of the things, I actually think that we're not formal enough about thinking about the -- this is a step forward from where we've been. We're still not formal enough in the

families of measures thing.

So, we'll talk in a second about the suite of tobacco measures taken together.

But I'd sort of like to do that periodically on everything. I'd sort of like to -- this is the suite of things that we current have for schizophrenia, what's not essential, what's missing? You know?

And you could do that with -- you can't do it with everything all the time, but I'd like to see us do that periodically in some kind of a formal way. And maybe you can map it to the continuum of care or something.

But you know, this is what we have in prevention. This is what we have in initiation. This is what we have in follow up. You know? I suspect that'd be really instructive.

CO-CHAIR PINCUS: So, Raquel and then Tami?

MEMBER MAZON JEFFERS: I actually think Peter just said exactly what I was going to say, that you could take Shane's idea of a roadmap

along the continuum of care and the various care settings, put the measures through that -- those different lenses and then do some serious housekeeping and get rid of the measures that are no longer relevant or not as strong as some new measures that have come in and really identify the gaps.

And I would go from like prevention to access to early identification, initiation, engagement, retention, maintenance, outcomes and then these recovery support measures which are housing that you're talking about.

So, I think it's really important, though, that the rubric that Shane was talking about be developed in a way first to help us place the measures in the context of the whole.

CO-CHAIR PINCUS: Tami?

MEMBER MAKE: I'm concerned that we don't have enough focus on the potential harms for these measures. We're -- almost all our measures are telling the system to do more stuff and we always assume that that's going to be

innocuous and then when the measures come back for maintenance, we have no data on the harms.

The harms could even be just more out of pocket spending. But, I think about how we got into the opioid epidemic, it was because we started getting an outcome measure about pain which we felt was a great idea and we ended up with this huge opioid epidemic.

So, I'd like to see more data coming back on potential harms. So, because I'm kind of, you know --

CO-CHAIR PINCUS: Okay.

MEMBER MARK: When I -- yes.

CO-CHAIR PINCUS: First do no harm.

Mady, did you --

So, let me try to summarize.

MS. MUNTHALI: I'm now going -- trying to go back to a number of comments. This has been very helpful.

And piggybacking on your comments about harm and going back to some original comments about how we could re-look at maintenance

measures.

When we looked at reassessing maintenance measures, one of the things we were trying to do is to reduce the burden on committees.

So, not to revisit some of the things that have passed before in terms of the evidence, if the evidence hasn't changed or the testing. And so, there is a great emphasis on things like feasibility and use and usability deliberately so we can get this feedback back from those that are using the measures, those that are being measured by it.

So, there is a global initiative for us to -- in order for us to reduce the burden, get rid of the measures that don't really matter, we need to know how they're doing that in the field. And that's something we haven't been focused on until very recently.

With regards to -- I'm kind of putting together a roadmap or -- Karen and I were talking and it sounded very much like what we used to do

in the past, haven't done very recently, but preferred practices for behavioral health. And we endorse those. And we can do that. That is definitely within the scope of this work.

One of the other things that we have done, you know, we are predominantly funded by CMS. It is unpredictable what projects we will get. So, we are not able to give advance notice to you or to developers about the projects that may be available.

But, one of the things that we did build into our contracting is what we call off-cycle review where, if a project isn't funding during the maintenance process, you can talk about things like a framework for your gaps, where you'd like behavioral health to go.

A really good example of how this has been operationalized is through our care coordination project. They were able to do this and they've built a framework around care coordination.

And I would encourage us, including

myself on the Behavioral Health Committee, but encourage the Behavioral Health Committee to think about that.

It will give developers, if this is taken to the incubator, more specificity. This is where developers come back to us with feedback. They get general comments about the need for maybe measures around schizophrenia. But what should they be focused on?

And, so, I think the more that you can get granular, the more that you can scope out a continuum of care, treatment, whatever, that gives them specificity and direction, the more we will be better able to get the measures that we really want.

CO-CHAIR PINCUS: So, let me add a couple of things and also try to summarize.

So, one thing that I thought was very notable in terms of when we looked at the different areas, that the most prevalent mental disorders are anxiety disorders, and that's not represented anywhere. And, so, that's one area

that clearly comes up.

A second area is, you didn't look at it in terms of children and adolescents. But, that's another sort of cut to think about. And here's sort of the overlap with the pediatric group and how to deal with that.

Number three, which goes, I think a little bit to really the last comment, that, you know, Lisa, you were making, which is, you know, take, you know, developing some kind of framework.

And it's related to, I think the thing that Shane brought up about, you know, a framework or some kind of recommended best practice.

But I think there were two different things there. One was really, I think, you were aiming it at, because you're a behavioral health provider, you know, what are the key recommended measures that you would use that would sort of give a, you know, a first level cut at like how well you're doing which is a little bit different

than sort of an overall framework for measure development which is something.

And I think -- so, I agree with -- but I think it's both and. I think that it would be useful to do that. But I also think it might be useful to convene us as a group to actually think about developing a framework for how we ought to think about best practices for measurement across this.

And in addition -- so, that's I think three or four items that I mentioned. There's another additional thought that I had which is that I think also it would be a good idea for you to convene measure developers, but in a more formal way.

Specifically around the topics we've been discussing today. Because, it's quite clear that, among the measure developers that presented, they don't talk to each other and they don't talk to other people about what are best practices. You know?

And I was surprised at how sort of

isolated, you know, those, you know, they are. And, so, I think that that's, you know, a responsibility for NQF to, you know, to think about how to solve that.

And, but I do think convening this group, you know, not necessarily to measure -- to review measures that have been submitted, but to actually come up with, you know, several different frameworks and revising some of the process and looking over in a formal way that was suites.

Just in the -- and then, to summarize the other things that have come up, one is also to look at something like this, but also looking at, you know, by use cases or by sort of accountable entities how the measures are getting stacked up to see that we have enough across the board.

I think the idea of a new maintenance process and form that has more emphasis, maybe it's okay if less emphasis is on the evidence stuff evidence of importance to measure unless

there's new evidence, but to actually look at what's happened as a result of the measures, how it's being used, what have been potential negative outcomes. So, that's really the focus and the expectation.

The idea of adding sort of allowing sort of stratification to be done without calling it a separate measure, but actually building it in to the standardized use of different measures that can be adapted. So, it's not like adding necessarily more burden.

The idea of nesting certain composite measures together in terms of screening, follow up, you know, consistent -- continued measurements to outcomes. You know, as, you know, like the, you know, the Minnesota Community Measurement measures that have been developed.

Next one is, you know, I think the approach to looking at non-health settings, quote, non-health settings like jails, housing, things like that that would incorporate that. I think that's a really significant gap.

The issue of access is huge. I mean, and there are, to my knowledge, really no measures of access to behavioral health services. And that is a major national crisis.

Sixty percent of psychiatrists don't take any insurance.

MEMBER MARK: Maybe NQF could clarify that because my understanding is NQF doesn't approve access measures.

So, for example, the identification measure which basically says, what percentage of people are getting STD treatment is a HEDIS measure, but it is NCQA people told me it can't be NQF endorsed because it's an access measure.

MS. MUNTHALI: That's not correct. We actually -- we do see a lot of access to care measures through our Health and Well-Being Portfolio.

And there was a struggle for measures to come through, quite frankly, because developers had a difficult time putting them together.

And, so, what that committee did is

develop a framework for developers to assist them so that they would know what kinds of things we should be looking for but also for the committee and everyone else who is looking at those measures. How should we be looking at access to care measures?

CO-CHAIR PINCUS: Could we see that?

MS. MUNTHALI: This framework -- we'll send you the framework. We just developed it last year.

CO-CHAIR PINCUS: Yes, but it's --

MEMBER MARK: But, that's really important.

CO-CHAIR PINCUS: You know, if you think of one area of healthcare where there's definitely an access problem, if you look at any primary care provider in the country, it's behavioral health. And there's no access measure.

Next item that would be a focus on outcomes. And looking at what are the expectations for best practices on measuring

outcomes? Which relates, obviously, to that nesting kind of issue.

But, you know, that needs to be sort of laid out and then communicated to measure developers in that regard.

You know, I think Mady mentioned before the issue of integrating mental health and substance abuse. So, it was Connie, both of you did.

I think that that's, you know, that clearly is also a gap. There's nothing that I could see that currently exists in that area in terms of coordination and linkage.

MEMBER CHALK: That main outcome measures may need a different path.

CO-CHAIR PINCUS: Right. And then different ways to capture them.

MEMBER CHALK: Yes.

CO-CHAIR PINCUS: I mean, like, you know, in a sense, you know, segmentation of, you know, preventive care measures for people with schizophrenia is kind of a coordination measure

as much as it is a process and potentially, you know, control of hypertension is an outcome measure, but it's also a coordination measure --

MEMBER CHALK: Right.

CO-CHAIR PINCUS: -- you know, in terms of those things.

Because, if you have shared accountability, then you're, you know, people have to talk to each other. You know, so those kind of things.

And then coming back to what was discussed at the end, putting together for us sort of what's the suite of measures as we're about to do now with regard to tobacco, I think, is a very useful exercise for us to do and to see where the gaps are.

So, let me turn it over to you.

MEMBER PINDOLIA: When you were talking, I thought of what Tami had mentioned and then, you know, when we do new initiatives or when we do a P&T review or launch a program, the first year is the most critical time to figure out what

are all the things you didn't think of when you develop.

But, when we do an approval, it's a three-year approval. Right? Shouldn't the first time, could we change it that the first time is a one-year approval so we can catch any patient harm issues or, you know, any inadvertent concerns that come up from that data so that we can address it early.

And I'll tell you like just from the CMS and a couple of the measures that we have to do and without people realizing, because you want them to get a five star and you're getting a one-time bisphosphonate for a senior and they really shouldn't be taking it when they're 79. But, you know, things like that.

MS. JOHNSON: You know, it's a great idea. I think a couple of things we do ask about, you know, any evidence of harms. As a matter of fact, that's one of the things specifically that we ask about under usability and use.

Usually, developers only tell us about what maybe they found in testing. They don't always tell us about what happens in the world. And, sometimes, they don't know, which is a problem.

So, what we've been trying to figure out is how we get people who do know to tell us. We haven't quite figured that out yet. So, if we can -- if you guys have any great ideas about how we can get that feedback, you know, have said for years, please tell us. But it's a passive please tell us.

(Off-microphone comment)

(Laughter.)

MEMBER MUNTHALI: Yes, right.

MEMBER COLEMAN: Yes, even a two-part application process where they come here first, we can at least generate which worries or questions we have and then they have a time and period to go investigate those of sorts before they come back or something like that.

I mean, something like that might at

least start to get at the spirit of some process like that.

MS. MUNTHALI: Yes, what we really -- we thought it was very valuable the ECHO discussion. We would like to do more of that before measures are endorsed so that you can see these measures before the developer goes further down the track of measure development.

And it looked like it was beneficial for you as well. I think for the developer as well so that they can change gears if they need to and we can get an early look at what might be coming down the pike.

MEMBER TRANGLE: Yes, you know, how do I say this? I think even what's in this group, but I'm think that Mike is probably one of these people. Some of the stuff that you're talking about that we should get together and maybe think about talking about, some of us have been playing with and developing and using this for quite a while. You know, it's imperfect.

We have an access measure, you know? It

measures outpatient access. It measures access to clinics. It measures -- you can count virtual consults and other kinds of access.

You know, we've got a composite bundle for schizophrenia, a suite of schizophrenia measures.

We're not a developer. It wouldn't occur to me to take time, effort and energy to bring it to you guys.

I think there's a fair amount of wisdom if we'd just sort of try to capture it. But, it hasn't been in the framework for you think about including that way. You know?

MEMBER LARDIERI: I just have a question. Does NQF have any idea where the measures get implemented, like which state, which health plans say, you know, we're going to implement this, you have to do this now? Is there any requirement on the developer to track where their measure gets implemented?

Because if there was some of that, then you might be able to get some of the data

quicker, faster and maybe that needs to be some place.

MS. MUNTHALI: It's difficult for developers, too, to know who and where our measures are implemented. There's a lot of misapplication of measures.

You may approve a measure for a certain level of analysis and it may be applied at a different level of analysis. We've seen that happen a lot.

We just did a project on variation of measurements specs. And not because people are, you know, meaning to do bad, but they're just trying to get around some of the challenges we mentioned before, some of the data challenges, some of the implementation challenges around collection and reporting.

So, the answer to your question is, no, although measure developers do have a contract with us, it's a measure steward agreement. We tell them to tell us when there are any updates, tell us when there are changes.

And, you know, it's part of their requirement. But, it is hard to enforce.

CO-CHAIR BRISS: And the addendum to that is that, every time I've been in a conversation along that line, it's like the conversation this morning where, you know, three people pop up and say, I know where your measure's been being used in these places and he said I have no idea. Right?

And, so, there's a whole lot of that and it's trying to get a better sense of where things are being used and if something good or bad happened is sort of an undiscovered country, in my view, that's gotten not nearly enough attention.

MEMBER ZIMA: I just wanted one comment, and that is that, I think with behavioral health more than probably other fields, we are more interested in how these quality measures are used in other sectors.

And we talked about jails and homelessness, but particularly for children, we

also think about health circuit med treatment and foster care and things like that, and in schools.

And, you know, and I just sort of wonder with NQF, are you taking the lead in advocating for these siloed data infrastructures to link?

MS. MUNTHALI: We are. We just endorsed our first home and community based services measure which was -- which came after a project on how we look at home and community based services and measures. So, we built a measurement framework around that.

And we would like to see measures that are specific to populations, have broad application. But it's really difficult, the data issues that we find, especially as you're trying to get into settings and areas in which are not -- you're going further away from the clinical settings and into populations that are marginalized vulnerable populations in which, for you to get a comprehensive look at, I mean, health, health care, quality of health and health care, you'd have to have a very integrated

approach.

What we find, and this is happening even as we're thinking about SDS, the biggest challenges are around data, linking the data. And, so, that is one of the things. It's not the thing that draws people, but it is -- it's a huge obstacle for us and for developers as well.

MEMBER LARDIERI: I have a real quick idea, why don't NQF set up an NQF registry for, you know, a voluntary registry so all those health plans state that use an NQF thing, go into your website, identify what they're using and if they're using the measure as is or if they've modified it.

And it wouldn't be that much of a heavy lift that way.

MS. MUNTHALI: So, that was almost a recommendation that came from our variation in measurement specs project.

One of the things that they were saying is, okay, we know that variation is happening. It's happening for a number of reasons, how can

we mitigate it?

At a minimum, we should be transparent. And the recommendation from the committee is that a body like NQF house repository or a database in which you can, you know, kind of catalogue all of this information.

So, that's something that we're looking into. We're hoping to, as with the attribution project, get additional funding to look into that and see if not us, who? And how would we do it?

MEMBER TRANGLE: I just want to make one last comment. My organization is part of NIMH got together and formed a mental health research network of systems that had reasonably large integrated multispecialty groups, inpatient, outpatient and health plans where we could look at the data.

And when that started, that network was maybe ten, and it's up to maybe 16 places now. It started out that all of our projects were the pet projects of researchers who had already gotten NIMH grants.

And over the last, I don't know how many years we've been around, maybe six, seven years, it's really evolved such that the questions that are being researched are coming from a delivery systems and the health plans that are part of it saying, these are burning issues for us to improve the quality of care or whatever it is that we deliver.

And we've also come up with certain kind of operating principles that, you know, if it's going to burden staff at a clinic or clinicians or secretaries or whatever it might be, it's not going to happen.

It's got to be back office, you know. It's got to potentially be sustainable. So, if it's an improvement, we might continue it.

But, and maybe we're at this same sort of evolution where, you know, developers who are in it for the glory or the money or both come to you with little slices of, I'm interested in this, I'm interested in that, with really no guidance as to whether it's a puny little thing

or it's a big ticket clinical issue.

And I hope we're evolving in the same way that this research network is evolving where it's got to be really by the -- and we may have -- may not be the right -- we may not be constituted as a committee to do this -- the right committee to give you that feedback.

But that's where the direction should be coming from and then the developers should be told this is what we need. You know, and if it's a n RFP process or whatever it might be instead of vice versa.

CO-CHAIR BRISS: Let me take a minute and share with you that the portfolio or most of the portfolio of tobacco related measures.

So, as you've heard me say this before that I'd love to be able to do this kind of an exercise periodically.

DR. LUSTIG: And I just wanted to point out, I tried to not try to shove everything onto the slide because too many columns and rows. So, you also have a handout that gives just a little

bit more detail.

CO-CHAIR BRISS: Yes, as this stands, you really have to be an aficionado to be able to make sense of it. But this shows most of the portfolio of tobacco related stuff. I'd love us to be able to periodically look at the suite of measures related to certain topics every now and then actually regardless of which committee, you know.

And, so, one of these actually came out of the Child Health or Pediatric Committee, whatever you call them. And --

DR. LUSTIG: Yes, and so, actually, the last column, we can now disregard because that's the measure from yesterday that did not pass.

CO-CHAIR PINCUS: I don't know, but I think that's where the --

CO-CHAIR BRISS: But they'll come back.

DR. LUSTIG: Yes.

CO-CHAIR BRISS: They'll come back.

DR. LUSTIG: Yes, it's good there.

CO-CHAIR PINCUS: I am disappointed, I

think it's actually important to think of it in this.

CO-CHAIR BRISS: Yes.

DR. LUSTIG: So, but the other -- the 2803 and 2020 are from -- are in other committees endorsed these.

CO-CHAIR BRISS: Yes, and so, if you think about the portfolio, taken together, there are sort of several measures that are about screening and/or counseling at the clinic or provider level.

So, 3225 and 3185 are the ones we reviewed yesterday that are sort of the -- and I've been doing these too long because I want to characterize these as the ones that used to be NQF 28, right?

The 2803 is a similar measure that's a pediatric measure. It's a little bit different on the treatment side because of the med treatment isn't recommended for kids.

The -- you saw NQF 27 yesterday as well. That's sort of a similar measure about screening

and treatment at the health plan level.

The 1651 to 1656 are three -- a suite of three related joint commission measures that are sort of did hospital screen and appropriately treat among hospitalized people.

2600 is one of those did you do the right thing in the behavioral health population to, right, that Mady was talking about.

The 2020 and 3229 are more population prevalence measures. They're not -- I wouldn't really characterize either of them as screening measures.

But 2029 is prevalence at the level of a patient panel that you saw yesterday. And 2020 is at the state level from BRFSS.

And, so, in this subject matter, it looks like a lot of measures. I can make them, at least in my head, to look like a coherent set.

So, you've got did you appropriately screen and treat in adults? Did you appropriately screen and treat in kids? Did you appropriately screen and treat in hospital

settings? And did you move the needle on prevalence in patient panels or states? Right?

And, so, but we should -- and treat, as I don't think that there are huge gaps in this portfolio, but we ought to be looking at portfolios as a whole like this.

And, you know, we talked about some ways yesterday that 3229, for example, might be consolidated with some of the processing measures.

You know, there are ways that some of these might be further composited or at least looked at together as a suite in a way that would communicate better.

So, this is an example of a set of measures. And this has already been said that they didn't all come out of our committee. And, so, it's good to look at the whole bunch of them together periodically I think.

MEMBER MAZON JEFFERS: So, I don't know if you're asking for this feedback, but there seems to be a tremendous amount of overlap and it

would -- a great opportunity for simplification with a little careful study.

CO-CHAIR PINCUS: So, Peter, I need some of your thoughts about what the opportunities are there for consolidation. You know, obviously, we have, you know, an alignment.

I mean, you know, one is getting, you know, some of the age issues, you know, sort of clear. Some of it might be, you know, you know, having sort of similar, if we're applying the concept of what is counseling to do it in a consistent way across, you know, is screening defined in the same way across all of these?

That includes 28 -- so, those are the kind of things that -- so, part of alignment, I mean, if you want to operationalize the concept of alignment, part of it is, you know, sort of nesting or consolidating very similar overlapping measures.

MEMBER MAZON JEFFERS: Or could a single measure --

CO-CHAIR PINCUS: Let me just finish.

MEMBER MAZON JEFFERS: Sorry.

CO-CHAIR PINCUS: And the second part is for similar concepts to use the same operationalized definition for those concepts.

And then if there's competing measures, to select the ones that are, you know, are the -- have the greatest feasibility, that don't, you know, in terms of how much effort is required to get it.

CO-CHAIR BRISS: So, Raquel, did you get --

MEMBER MAZON JEFFERS: The only other thing I would add is -- or if there's a single measure that we really like how it's constructed, could it simply be applied in different settings?

CO-CHAIR BRISS: So, I'm personally less -- so, for example, I don't think that -- I think that counseling and treatment and primary care and counseling and treatment in hospital settings, for example, are both important and I personally wouldn't favor dropping on in favor of the other, at least for something that's as

important as tobacco.

For example, there are people -- there have been people at CMS that feel like one of the -- like everybody could measure one of the outcome measures. And we could drop all the process measures.

I personally am skeptical about that, truth is. You know, but that's the conversation that kind of -- at this point, I'm not trying to give a formal vote to how I would solve this mess.

I think reasonable people could -- at a minimum, I think we could do better about presenting and talking about our suite of things and what they -- what we think they uniquely contribute.

So, in this one, I can make a case for unique contributions maybe with the exception of I don't much favor -- I think I'm echoing Mady a little, I don't much favor did you do appropriate chronic disease prevention in a special -- in a population with some other comorbidity because

I'm afraid that what that always gets you is 50 measures.

You know, did you do smoking prevention in people with mental health makes perfect sense until you also have to do it for people with diabetes and cardiovascular disease and every other condition under the sun.

And, so, I don't actually favor -- I don't much favor measures like 2600 on that basis. The rest of these, I can make a case for that you could -- the joint -- there's three joint commission measures could be composited in principle instead of having three measures that are essentially about did you screen? Did you treat it in the hospital and did you assure treatment going out the door?

It could be a smaller number of measures. You know? And, so, there are opportunities to talk about that stuff. But, I do think lining these up and having that conversation is a good thing.

And the point of this at this moment

isn't to micro manage that conversation but to talk about -- I think we ought to do more of this and have more of that conversation.

CO-CHAIR PINCUS: But, I guess, just to follow it up in terms of the -- but, what is the next step? I mean, we can opine that, you know, that's there possibilities of X, Y and Z. But, who -- how does that get operationalized into actually a change?

What's the role of NQF in, you know, communicating to measure developers or the groups commissioning measure developers like CMS?

I mean, is it, you know, is it our -- is it, you know, a responsibility of this committee, obviously, at another meeting, to you know, to say, well, here, why don't you consolidate, you know, these three.

Add a, you know, a secondary screen X amount of time after, see how many people stopped smoking. You know, I mean, you know, is it our job to do that or is it, here's issues, the field figure it out.

CO-CHAIR BRISS: I have a couple of suggestions. So, a couple of suggestions is that I generally think that when a committee gets to this point in a meeting and everybody's exhausted, we all shoot from the hip and opine about what are my three favorite things that I'd like to see additionally done? And I think that that's wrong.

So, I'd love to see us, even if we could just tee up a couple of subject matters every once in a while, I'd love to see us tee up something like this on a periodic basis, point one.

And then we could at least give some more informed feedback about what we think we can do more of and what we'd like to do less of is point one.

Point two is that you guys are always looking for constructive stuff to do off-cycle. Right? And I think we could do this kind of exercise in off years.

So, imagine a state of the world where

we did a -- say, for the purposes of argument, next year or the year after is an off-cycle year for the behavioral health committee.

I could image we did a quarterly webinar and teed up one of these things every quarter. And we could get a lot of constructive stuff done even in an off year, I think. Right?

And, yes, and it would have the additional advantage of having committees stay at least a little in contact and not have to feel like they're completely starting over in the years where they do get funded. You know?

And, so, at least some of that kind of stuff could get done and the results could be -- the results of that could be used both to pair the portfolio, which I'd really like to see done, truth is, and to feed the measure developers and the incubator and other things that might have opportunities to fill gaps or -- and/or improve existing measures.

MEMBER KELLEHER: I second that.

CO-CHAIR BRISS: Wow, I've thirded. I

need to quit while I'm ahead.

Anybody else have comments about this kind of stuff?

MEMBER CHALK: The only comment I wanted to make is that, the population prevalence measure, the 2020 I guess it is, there could be some conversation about how population prevalence measures are used to nest everything else that we're talking about.

In the area of substance use disorders, we're talking about those measures being contextual measures and being essential to everything else that's built on top of them.

You want to know what the prevalence is and then you build on top of that a bunch of measures.

They set the context for everything else.

CO-CHAIR BRISS: Yes, and maybe could I add to that? I mean, one of the problems with those kind of measures is we didn't have as much conversation about this as I thought we might

yesterday.

But, there's -- particularly, these kind of measures have -- the outcome measures have a big attribution problem. Right?

And there's inevitably a lot of worry about penalizing providers that take care of high risk populations. Right?

And if you -- and these measures would be easier to sell, I think, if they were always packaged with other things and weren't treated as if they were standalone and going to be, in my view, and if they weren't packaged as if they might be standalone high stakes measures that we're going to penalize people that we're taking care of sick people. Right?

Yes, Michael?

MEMBER TRANGLE: Your suggestion of eliminating, which I get when you start generalizing it, but eliminating looking at how we're doing this measure with patients with serious mental illnesses, I get that you don't want to do it in every population.

But, on the other hand, if you truly eliminate that without having this a dispirit group or at least a sub-specification --

CO-CHAIR PINCUS: And I think the assumption is that one would have a generic measure and one could stratify that.

MEMBER TRANGLE: And, so, mandating stratifying so that we can see within that population then it sounds fine. Otherwise, we're back to where we are now where it's lost.

CO-CHAIR BRISS: Although, like NCQA, for example, today does things like some years they mandate that you report your performance on 27 in a specified subpopulation. Right?

You know, so, they have years when HEDIS reporting says we want to focus on -- we want to have a focus on mental health or cardiovascular disease or whatever it is.

And, so, there's capacity in some of these today to do something that's kind of similar to what 2600 is trying to do. I think we could do that in other subjects.

MEMBER TRANGLE: But make it a ten-year period, not this year.

CO-CHAIR BRISS: Yes.

CO-CHAIR PINCUS: No, but I think that if you -- the issue is stratification is that you can do it and then that gives, you know, as a single measure and it gives options for stratification but do it in a standardized way so everybody's using the same methodology for defining a stratum.

CO-CHAIR BRISS: And the other thing that we need to keep in mind is that it's not going to be generally true that all these have -- all of these collect -- currently collect the data that would allow you to stratify it. So, we'll have to be careful about -- that raises its own set of burden issues, too.

So, we're going to have to be -- that's -- there's never a perfect measurement solution. Right? That's got no cost and no harms, but at least that's something that can be talked about.

MEMBER TRANGLE: It might be an

incubator kind of issue to work on. Which ones could be stratified to get mental health? I can see the excitement reigning supreme over there.

(Laughter.)

CO-CHAIR BRISS: All right, Tami?

MEMBER MARK: You're on a winning streak so I had to disagree with something.

CO-CHAIR BRISS: Okay, good.

MEMBER MARK: Yes, I guess I get the point that it gets too onerous when you start slicing and dicing. But, most people who are going to pick up these measures aren't going to know that there's a huge problem with smoking among SMI or aren't going to know that there's a huge problem with cardiac conditions among SMI unless there's a separate measure.

So, that's been my experience.

CO-CHAIR BRISS: I agree.

MEMBER MARK: So, that's the counter-argument.

CO-CHAIR BRISS: Mike?

MEMBER LARDIERI: I'm trying to say,

Peter, I think, you know, if we move to eMeasures and I don't think then the burden gets that -- to stratification gets that difficult. Because you already have the data in there and you're going to be able to stratify whichever, you know, whichever component or silo that you're looking at as it gets less difficult.

CO-CHAIR BRISS: It does, assuming that you can find the data in an appropriate field. You know, there's never a perfect solution as I think I -- my answer to everything at this point is there's never a perfect solution. Right?

MEMBER MAZON JEFFERS: Would be okay to take five minutes right now and have someone explain to those less initiated, what is the measure development process and where do the costs sit?

The question that Mike asked earlier, I feel like it would help us understand contextually the cleanup process as well.

(Laughter.)

CO-CHAIR PINCUS: We have seven minutes

because we have to stop at three.

(Off-microphone comments.)

(Laughter.)

MS. MUNTHALI: So, we actually could do that. Actually, we have -- we do engage with measure development -- measure developers outside of the review. We do have a standing measure developer committee, but it's a community.

We meet with them monthly. We do webinars, but we also have an advisory group of measure developers.

And, as through the incubator, we have a learning collaborative.

So, those issues I think, Harold, you talked about, you know, noticing that developers are not talking to each other. It's -- that is very true. Because they do have different business agendas.

These are organizations that are developing measures, you know, for the greater good, but also they have business reasons for why they're developing measures.

So, it is very difficult. It's something we don't talk about often, but it is a reality. And it's very expensive. We don't know -- we have rough figures around how much, you know, it may cost, but we don't know because some of that is proprietary.

We see those measures once they are close to be fully spec'd or spec'd here. But, it is a very, very long process that involves testing.

A lot of the obstacles that they encounter is around the testing. They go into something and sometimes things don't work out the way they anticipated. Their access to data, their -- we can have a developer walk you through that.

We went through this when we did a lean improvement activity around our consensus development process, a kaizen, in which we had a stream that looked at measure development from concept on to testing. And we were like, wow. We were quite impressed.

There are a lot of steps, a lot of challenges, things that go into the, you know, into the pipeline don't always come out.

And, so, it will be very difficult. That's why we have been stressing since that kaizen to really do a lot more upstream engagement with developers before they even start thinking about ideas for developing measures, really being able to influence that there as opposed to when they come to you here.

Because a lot has gone into in terms of resources, human, all sorts of resources.

CO-CHAIR BRISS: And the only other thing I would add is that I don't get the sense that this is much of a money center for anybody, it's the generally feeling in the field seems to be that it's a long-term expensive project to get any measure to the table.

It's probably on the order of a quarter of a million dollars and years to get any measure here which is why they get so -- one of the reasons they get so gray. These people have --

when they get to here and we turn them down, they've just sunk a quarter of a million dollars and two and half years.

And there's not much of a financial reward if they get through.

So, I think it's -- I personally think it's too much of a bottom up process and I personally think that it's got all kinds of problems. I don't think that there's -- that the primary problem is that it's a lucrative thing to be a measure developer.

I do agree that having too many measures and especially too many low value measures generates a lot of costs to the system that are really important that nobody thinks about enough at any level of the measure development process which is one of the reasons that you've heard me on my soapbox constantly about I'd like to see less measures. I'd like to see less measures. That's what I'm thinking about.

But, I don't think a primary problem is likely to be very --

DR. LUSTIG: We technically have three minutes left and we actually need to ask about public comment and Kirsten has some follow up.

MEMBER LARDIERI: Okay, so maybe it's in a different discussion, but maybe we would be a little more transparent because nobody's putting a quarter of a million dollars in there for nothing. For all the measures that we have, that's not happening. So, maybe we need to be a little more transparent about that.

I know there's -- do they pay NQF a fee to get this? So they don't do that. But, you know, there's more than, you know, nothing to it.

CO-CHAIR PINCUS: Well, you need to distinguish between measure commissioners and measure developers. There are groups, most commonly CMS, that commissions other groups to develop measures, you know, that as their clients.

CO-CHAIR BRISS: So, we've got -- the staff are trying to get us to public comment.

Not, right now, we can further this

conversation in a follow up phone call if you guys want to. So, can -- would anybody on -- either on the phone or in the room like to make a public comment?

OPERATOR: And, at this time, if you would like to make a public comment on the phone line, please press star one.

And there are no public comments from the phone line.

CO-CHAIR BRISS: Okay, follow ups? Follow ups?

DR. LUSTIG: So, one of the questions is we do have reserved time next Thursday for a post meeting call. We do that in case there are any issues related to the measure endorsement that weren't resolved. We don't have anything related to that to discuss, so we can cancel the call or we can hold it and continue to talk.

But, otherwise, we can go ahead and cancel that and then we can prepare some things related to our discussions today and have that for a future time.

CO-CHAIR BRISS: And before everybody sort of goes off to the four winds, I want to thank the staff for all their ages and ages of prep work.

(Applause.)

CO-CHAIR BRISS: And especially this stuff which was mostly new people who hadn't done this before. So, congratulations on surviving your first one.

(Off-microphone comment.)

DR. LUSTIG: Next Thursday, yes, it sounds like we're going to cancel that.

CO-CHAIR PINCUS: And I also want to thank all of you because, I mean, we really had some superb, vigorous discussion. People really had great input and that's the kind of, you know, work that we like to get done at these meetings.

DR. LUSTIG: And, so, before you run off, Kirsten, you just want to talk about, just so you know the process after this.

MS. REED: Sure. Okay, so, this March 9th webinar, we will go ahead and cancel. So,

enjoy your free two hours next Thursday.

In the meantime, Tracy, Desi and I are going to be drafting the report to prepare for the public commenting period which will run through April 5th through May 4th.

And then, once we do get all of those comments, we will compile them and prepare for a post draft report comment webinar with all of you where we will present all of the comments we received and start preparing our responses to those.

That does need to be scheduled, so I will reach out and see which dates and times work best for you. But, it will be a couple weeks after that May 4th deadline or ending of the commenting period.

Other than that, thank you all for coming. I think our meetings department will be following up with you with the reimbursement form for all of your expenses in traveling here. If you don't get that or have any questions about it, please let me know and I will talk to you

soon.

DR. LUSTIG: And everyone got thanked except our co-chairs. So, I'd like to thank our co-chairs for doing a great job.

(Applause.)

(Whereupon, the above-entitled matter went off the record at 3:01 p.m.)

M

T

W

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 Standing Committee

Before: NQF

Date: 03-01-17

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R. Gross

Court Reporter