National Quality Forum
Primary Care and Chronic Illness Standing Committee
Primary Care Measure Evaluation Web Meeting Fall 2021 Cycle
Friday, February 11, 2022

The Committee met via Videoconference, at 10:00 a.m. EST, Dale Bratzler and Adam Thompson, Co-Chairs, presiding.

Present:

Dale Bratzler, DO, MPH, OU Health; Co-Chair
Adam Thompson, BA, Northeast/Caribbean AIDS Education & Training Center; Co-Chair
Kim Elliott, PhD, Health Services Advisory Group
William Glomb, MD, FCCP, FAAP, Superior HealthPlan
David Lang, MD, Cleveland Clinic
Esther Babady, PhD, D(ABBM), Memorial Sloan Kettering Cancer Center
Katherine Gray, PhD, formerly with Sage Health Management Solutions
Lindsay Botsford, MD, Iora Primary Care and One Medical
William Curry, MD, Penn State College of Medicine
Anna McCollister, Galileo Analytics
Harold Pincus, MD, Columbia University
Jeffery Susman, MD, University of Texas Medical Branch
Vitka Eisen, MSW, EdD, Healthright 360
Raquel Mazon Jeffers, The Community Health Acceleration Partnership
Brooke Parish, MD, Health Care Service Corporation - Blue Cross Blue Shield of New Mexico
Karen Fields, MD, Moffitt Cancer Center
Shelley Fuld Nasso, MPP, National Coalition for Cancer Survivorship
Jette Hogenmiller, PhD, MN, APRN/ARNP, CDE, NTP,
TNCC, CEE, Oncology Nurse Practitioner
Jennifer Malin, MD, PhD, Optum Health Solutions

NQF Staff:

Paula Farrell, MSHQS, BSN, RN, CPHQ, LSSGB, Director
Poonam Bal, MSHA, Senior Director

Tricia Elliott, DHA, MBA, CPHQ, FNAHQ, Senior Managing Director
Oroma Igwe, MPH, Manager
Gabrielle Kyle-Lion, MPH, Analyst

Also Present:

Susannah M. Bernheim, MD, MHS, Yale Center for Outcomes Research and Evaluation
Greg Bocsi, DO, University of Colorado
Michael Jellinek, MD, Massachusetts General Hospital
Michael Murphy, EDD, Massachusetts General Hospital

Contents

Proceedings

(10:00 a.m.)

Welcome and Review of Meeting Objectives

Ms. Farrell: Hello, everyone, welcome, and thank you for joining the Primary Care and Chronic Illness Fall 2021 Measure Evaluation meeting. I'm Paula Farrell, the director of the project, and today we have three measures that we're going to be discussing.

This meeting is going to be a bit unique in that NQF consolidated some of its fall 2021 projects and reassigned the evaluation of certain measures to neighboring CDP portfolios. So today, we're going to be evaluating one primary care and chronic illness measure, one measure from the cancer portfolio, and one measure from the behavioral health portfolio.

To assist in that evaluation of those additional measures, we have invited Behavioral Health and Cancer Standing Committee members to this meeting. Everyone is considered a member of this committee and will have the opportunity to discuss and vote on all three of the measures that we evaluate today.

So now I just want to simply turn it over to our co-chairs, Dr. Dale Bratzler and Mr. Adam Thompson, to provide their welcoming remarks.

Co-Chair Bratzler: Anna, good morning, thank you. Dale Bratzler here, good to see some familiar faces again. We become friends over time on these standing committees, appreciate all the work that you've put into reviewing it. I took time to read all the comments that came in and really appreciate all the work that's gone into it, so look forward to the meeting.

Co-Chair Thompson: Good morning, everyone. I echo Dale's statements. I just want to say welcome also to

our partners from the Cancer and Behavioral Health Committee, to our Committee today, as well as our patient partners. We're really glad to have you here. And let's get started.

Ms. Farrell: All right, next slide, please. So next I'm just going to go through a few housekeeping reminders. And then we're going to start with introductions and go through disclosures of interest.

So we are on a Webex meeting this morning with audio and video capabilities. So we do ask that you please turn your video on if possible, because that just makes it so that it seems like we have more of an in-person conversation.

We also ask that you remember to please put yourself on mute when you're not speaking. And we encourage you to use some of the following features that are available on Webex. And there is a chat box that's available to you. And you can either message NQF staff individually or message the meeting attendees.

Using the chat is a good opportunity if you're just generally agreeing with a comment or if you have something that you'd want to share, you can go ahead and type that in to the chat. And we'll make sure that we either call on you or that we read what you've entered into the chat.

We do ask that you please raise your hand to be called on or use the raised hand function to be called on by the co-chairs instead of just speaking up. This allows us to ensure that everyone who wants to speak does have an opportunity to do so.

And finally, if you're experiencing any technical difficulties, please feel free to contact the NQF project team at primarycare@qualityforum.org.

Next slide, please. All right, so now I just want to introduce our staff that has helped us with our meeting today. I am Paula Farrell, the director for the

project. And next we have Oroma Igwe. She is the manager on the project. Then Gabby Kyle-Lion, she is our analyst. Adam Vidal is our project manager. Poonam Bal is our senior director, and Peter Amico has also joined us this morning. He is our consultant.

And Adam, Poonam, and Peter are supporting staff for the project. But we wanted to introduce them as they may be joining our group, joining in our meeting and all the discussions that we're having.

Next slide, please. All right, so our agenda for today includes introductions and disclosures of interests. And at that time we will ensure that we have quorum to hold the call. We're also going to provide an overview of our evaluation process and voting process. And then we're going to test out our voting just to ensure everyone has access and is able to vote.

So everyone that is on the Standing Committee, the Standing Committee members should have received an email from us with a voting link. And you will need that for this meeting so that we can test the voting, and then to vote on each measure later on. If you don't have that link, please let us know in the chat function or send an email to primarycare@qualityforum.org, and we'll get that link out to you.

After our voting test, I'm going to provide a brief introduction to the measures that we're going to review today. And then I'll hand it back over to our co-chairs to lead the discussion by the Standing Committee on our first measure.

We'll have about an hour that's planned for each measure. And NQF Measure Number 3667 is going to go first. We'll take a short lunch break at noon and reconvene around 12:30 Eastern Time to review the two additional measures.

We will also review any related and competing measures and will then end the meeting with NQF

member and public comments to see if they have any additional input to provide. And then we'll inform you of next steps and what to expect going forward.

Introductions and Disclosures of Interest

Ms. Farrell: Next slide, please. All right, so now I am going to turn the call over to our senior director, Poonam Bal, for Committee member introductions and disclosures of interest. Poonam?

Ms. Bal: Before I do, I just want to see if Dale or Adam wanted to do any opening remarks.

Co-Chair Bratzler: I don't have any additional remarks at this point.

Co-Chair Thompson: No, we're good.

Ms. Bal: All right. So I do want to thank everyone for their time. Today we will combine introductions with disclosures of interest. You received two disclosures of interest forms from us. One is our annual disclosure of interest, and the other is disclosures specific to the measures we are reviewing in this cycle.

In those forms, we've asked you a number of questions about your professional activities. So today we will to verbally disclose any information you provided on either of those forms that you believe is relevant to this committee. We are specifically interested in grants, research, or consulting related to this committee's work.

Just a few reminders, you sit on this group as an individual. You do not represent the interests of your employer or anyone who may have nominated you for this committee. We are interested in your disclosures of both paid and unpaid activities that are relevant to the work in front of you.

Finally, just because you disclose does not mean that you have a conflict of interest. We do verbal disclosures in the spirit of openness and

transparency.

Now, we'll go around our virtual table starting with your committee co-chairs. I'll call your name. Please state your name, what organization you are with, and if you have anything to disclose. If you do not have disclosures, please just state that I have nothing to disclose to keep us moving along.

If you experience trouble unmuting yourself, please raise your hand so our staff can assist you. And as Paula mentioned earlier, you can chat us or email us if you're having any sort of difficulty with the web platform.

All right. So with that, Dale Bratzler?

Co-Chair Bratzler: Yes, I'm here. I'm with the University of Oklahoma, and I have no disclosures.

Ms. Bal: Adam Thompson?

Co-Chair Thompson: Adam Thompson, patient partner with the Northeast/Caribbean AIDS Education & Training Center. And I have no disclosures.

Ms. Bal: Esther Babady?

Member Babady: Yes, I'm Esther Babady. I'm at Sloan Kettering Cancer Center in New York, and I have no disclosures.

Ms. Bal: Thank you. Lindsay Botsford?

Member Botsford: Good morning, Lindsay Botsford, family physician in Houston with Iora Primary Care and One Medical, no disclosures.

Ms. Bal: Kim Elliott.

Member Elliott: Hi, Kim Elliott. I work for Health Services Advisory Group, and I have nothing to disclose.

Ms. Bal: Thank you. David Lang?

(No response.)

Ms. Bal: Katherine Gray?

Member Lang: I'm sorry, I was muted. I represent the American Academy of Allergy, Asthma & Immunology. I work at the Cleveland Clinic. I have no disclosures relevant for the current work of this committee.

Ms. Bal: Perfect. Katherine Gray?

Member Gray: Yes. I'm in transition. I was formerly with Sage Health Management Solutions which is a victim of COVID. And I have no disclosures.

Ms. Bal: William Curry?

Member Curry: Hi, I'm a family physician at the Penn State University College of Medicine, the Department of Family and Community Medicine. I have no disclosures.

Ms. Bal: Perfect. William Glomb?

Member Glomb: Hi, I'm Brendle Glomb. I am a pediatric pulmonologist and neonatologist out of Austin, Texas, currently the senior medical director at Superior HealthPlan, the Centene Corporation, and I oversee our value-based contracting which gives me a little bit of extra insight as to how these measures are used down the road. Thank you.

Ms. Bal: Anna McCollister?

Member McCollister: Hi, Anna McCollister. I am a consultant that works on a number of issues related to patient empowerment. And I'm currently working on data with a genetic testing company.

Ms. Bal: Thank you. Harold Pincus?

Member Pincus: Hi. I'm Harold Pincus. I am a professor and vice-chair of psychology at Columbia University and also a senior scientist at the RAND

Corporation. I'm on advisory committees for Cerebral, Ableto, and Magellan. And I have grants from the Commonwealth Fund, the Foundation for Opioid Response Efforts, West Health Policy Institute, the John A. Hartford Foundation, and NIH.

Ms. Bal: Perfect. Thank you. Jeffery Susman?

Member Susman: Good morning. I'm another Behavioral Health Standing Committee representative. I'm a family physician, geriatrician. I just transitioned to the University of Texas Medical Branch in Galveston. And I have no disclosures.

Ms. Bal: Thank you. Vitka Eisen?

Member Eisen: Hi, my name is Eisen. I'm CEO for Healthright 360, a non-profit behavioral health and primary care provider for low income people in California. And I have no conflicts to disclose.

Ms. Bal: Thank you. Raquel Mazon Jeffers?

(No response.)

Ms. Bal: Brooke Parish?

(No response.)

Ms. Bal: Karen Fields?

Member Fields: I'm Karen Fields. I'm a physician at Moffitt Cancer Center in Tampa, Florida. And I have no disclosures related to these measures.

Ms. Bal: Thank you. Shelley Fuld Nasso?

(No response.)

Ms. Bal: Jette Hogenmiller?

(No response.)

Ms. Bal: Jennifer Malin?

(No response.)

Ms. Bal: Heidi Floyd?

(No response.)

Ms. Bal: All right, well, I'm getting some chats letting me know that we have certain members that are here but just didn't have a chance to speak up. So if your name was called, and you didn't have a chance to introduce yourself, please jump in now.

We just received a note from Raquel saying that she is on. But her voice capacity is not working. So our staff can work with her to resolve that.

Is anyone else on that would like to provide their introduction that has not had the opportunity to?

(Simultaneous speaking.)

Member Malin: Sorry. This is Jennifer Malin. I was having trouble getting in. But I'm one of the Cancer Committee standing members. My disclosures are that I'm an employee and have stock in UnitedHealth Group.

I'm not sure how much we're supposed to say in our introduction. I think I've been part of the Cancer Standing Committee for about ten years. My medical and college years, I spent the first half of my professional life developing quality measures and measuring outcomes for individuals with cancer and focused on palliative care. And I'm currently the chief medical officer for Optum Health Solutions.

Ms. Bal: Thank you. And I heard one more voice.

Member Parish: Hi, this is Brooke Parish. I was also having trouble with audio and some of the Internet connections. I came back from deployment and found they moved my office. So that was interesting this morning. I am with HCSC, and executive medical director, and on the Behavioral Health Standing Committee. Thank you.

Ms. Bal: Thank you. Anyone else that was not able to

introduce themselves that is on the call?

Member Jeffers: Wondering if you can hear me now. This is Raquel Mazon Jeffers.

Ms. Bal: Yes. You're a little faint, but we can hear you.

Member Jeffers: Okay, hi. So this is Raquel Mazon Jeffers. I am a senior director with the Community Health Acceleration Partnership which is a venture philanthropist group focused on safety and community health systems. And I have nothing to disclose.

Ms. Bal: Okay, perfect. Anyone else that is on that wants to do an introduction. I think we have a few individuals missing. I just want to pause one more second to see if there's anyone else.

Great. Well, thank you all for that. So I'd like to let you know that, if you believe that you might have a conflict of interest at any time during the meeting as topics are discussed, please speak up. So even if you've not disclosed it you can do so later on. So you can do it real time during this meeting or you can send a message or a chat to your chairs or anyone on the NQF staff.

If you believe a fellow committee member may have a conflict of interest, or is behaving in a biased manner, you may point this out during the meeting, send a message to chairs, or to NQF staff.

Before we move forward does anyone have any questions or anything that you would like to discuss based on disclosures made today?

Okay. And I did get a notification that Heidi has joined. Heidi, are you able to provide disclosures and provide the introduction as well?

(No response.)

Ms. Bal: Okay, we might have to come back to her. I

believe she just joined the call. All right. Then, Paula, I'm going to hand it back to you.

Member Fuld Nasso: Hi, it's Shelley Fuld Nasso. I just joined. And I am the CEO of National Coalition for Cancer Survivorship and co-chair with Dr. Fields of the Cancer Standing Committee. And I have no disclosures.

Ms. Bal: Perfect, thank you, Shelley. Great, so now I'll give it back to Paula. Thank you, Shelley, for giving your introduction.

Overview of Evaluation Process and Voting Process

Ms. Farrell: Great, thank you, Poona.

So I'm now going to turn things over to our project manager, Oroma Igwe, and she's going to provide an overview of our evaluation process and voting tasks. Oroma?

Ms. Igwe: Great, thank you, Paula. Good morning again, everyone. So at this time we're going to do a brief overview of the evaluation process. Many of you may be quite familiar, but it's always good to refresh our memories.

So your role as a Standing Committee member here is to act as a proxy for the NQF multiple stakeholder membership. As the GPC Committee, Geriatrics and Palliative Care Committee, you know, you not only generally oversee the geriatrics and palliative care measures, but you work collaboratively with NQF staff to evaluate measures against unfair criteria.

You provide a recommendation for endorsement, and your view, and your subsequently responsive comments that are submitted during our public comment period.

Next slide, please. So meeting ground rules, as you know, I want to remind you that this is a shared space of an interdisciplinary multi-stakeholder review. And we want to emphasize that every voice

is important. Our emphasis is that each committee member holds equal value on this call and also beyond in the vital scope of our work.

As members of the committee, you know, we do our due diligence to encourage you to prepare adequately prior to the evaluation. And we certainly invite you to today's meeting to remain actively engaged during the course of the call and also be cognizant of the varying experiences of those on the call.

Generally speaking, your evaluation will be based and grounded on our criteria and guidance. So hopefully, as you guys were preparing for this, you also referenced the guidance to which we do our evaluation.

Next slide, please. So this slide describes the process. It's a nice outline here. And this is the process by which we will conduct today's measure discussion and evaluation. Each measure discussion will begin with a brief introduction by the developer, and the facilitation will then be led by the co-chair.

The discussion will be stewarded by our assigned lead discussant and subsequently the supporting discussants. I just want to say thank you again for our discussants who will be leading in that regard today. Thank you so much.

The lead discussant will briefly explain information on the criterion, emphasize any notable areas of concern, pull out any also notable comments from the pre-evaluation survey that was taken by fellow Standing Committee members and, if needed, note the preliminary staff rating.

The developers will be available on the call to respond to questions. But that will be at the discretion of the Committee. So we want to maintain some level of order on the call. We will then, of course, open the full committee discussion, and then we'll move to a motion to vote on the criterion, before we go to the

next criterion.

Next slide, please. So here, many of you all know, is our endorsement criteria. And our endorsement criteria really guides the way in which we proceed to the measure evaluation. It's guided by five key criteria here, several of which carry a set of sub-criterion.

You know, please know you can see it mentioned here in bold font. Please note the must-pass nature of certain criterion, you know, is established here as ground rules.

But also note that there is some distinction on how we perceive the voting outcome depending on the measure status of either maintenance or new measure maintenance, meaning a returning measure.

Next slide. So voting on endorsement criteria, as written on the slide, is basically describing the way in which we will be voting today. This is the sequence.

Go to the next slide, yes, but I'm going to say the next slide will speak more in detail to the procedure. But generally, the process will be that we proceed through the criteria according to the guideline here. And if we do get to the conclusion of the vote, the outcome will be taken during the overall suitability for endorsement.

Next slide. So here is some detail on procedural notes. If a measure fails on one of the must-pass criteria, do you know that there is no further discussion or voting on the subsequent criteria for that measure? The committee will simply move on to the next measure. However, if consensus is not reached, then the discussion will continue.

However, the vote will be reserved for the post-comment meeting. And that also includes the overall suitability vote. That will not be taken on this call. And then the post-comment call is where we would

take the vote to move any criterion outside of that, since it's not recent.

The related and competing discussion note here is really a provision for a section of our call that will be held to just give you an overview of the related and competing measures. We certainly aren't taking any sort of vote or best in class motions on relating and competing measures. So it will simply be a discussion.

Next slide. So achieving consensus, that is what we're here to do. And this slide is very important. We like to break down the information so that you guys are aware. So 66 percent of active committee members is required to be a quorum to then vote live on this call. So I'm happy to say we are well above that.

We will do a tally again when we do our voting test so that we have our number right on. So just want you to know we will be taking a live vote on the measures today.

In the chart, you can see the percentage breakdown that will dictate whether it's pass, no pass, this is outreach, so on and so forth. So just keep these numbers and metrics in mind.

Yes votes are a total of the high and moderate based on the number of the active and voting eligible Standing Committee members once it is called.

And like I said earlier, measures which are not recommended, they will move on to public and NQF member commenting. But the committee will not re-vote on the measures during the post-comment meeting unless the committee decides to reconsider them based on submitted comments or former reconsideration requests.

So just know that if a measure is not recommended, is not moving forward, there is not necessarily a re-vote. But like I said, there is an opportunity for the

developer to submit a formal reconsideration and then give the Standing Committee another opportunity to assess it.

That is a little bit different from the CNR voting outcome that I mentioned earlier. Any measure, criterion that was from CNR will be voted upon again post-comment meeting.

Next slide. All right. So another mention on this very important, so far, committee quorum and voting. We may have, you know, variable attendance throughout the call. That is totally okay.

But if you find that you have to step away please, hopefully, use the chat feature and let us know. Or if you find an appropriate time on the call to open your line and let us know that, please let us know that.

And that will help us keep track of our numbers just in case we were to go below quorum especially, of course, we want to maintain 50 percent attendance for this call. Fifty percent is required to just even hold this call, 50 percent. But I guess you can say 51.

If we do not have quorum at any point during the meeting, the live voting will stop. And we will follow-up how we would proceed with that offline. But I believe don't leave, or we'll be in jeopardy of losing quorum today. So just some notes to keep in mind here.

Next slide. Great, so that really sums up the general process today. I will pause here for any questions from the group.

Okay, having not seen anything in the chat or hearing anything on the call, I will wrap up this portion and hand it off to my colleague, thank you all for joining.

Voting Test

Ms. Farrell: Thanks, Oroma. So now we're going to do our voting test. So all of the Primary Care and Chronic Illness Standing Committee members should

have received that link regarding the voting that I discussed earlier. So you'll need that at this point.

And I'm going to turn it over to Gabby so we can test that out to make sure everybody has access.

Ms. Kyle-Lion: Hi, everyone. Sorry, I'm just pulling up my screen. Give me one second.

All rightie, so like Paula said, we sent the voting link via email. If you do not have access to that link, please let us know, and we will send it to you. And again, just as a reminder, this is only for the Primary Care Standing Committee members.

Our test question today is do you like Brussels sprouts. Please select A for yes or B for no.

(Pause.)

Ms. Kyle-Lion: We are currently at 13 votes, and we need a minimum of 14 to continue.

Member Jeffers: Can you just reiterate if we should not be voting if we're not Standing Committee members, just to clarify?

Ms. Kyle-Lion: You are all -- you are a Standing Committee member. Yes, so you can vote. Yes.

Member Jeffers: Thank you.

Ms. Kyle-Lion: On the Primary Care, Behavioral Health, and Cancer Committee, I should have clarified. Sorry.

Member Jeffers: Yes.

Ms. Kyle-Lion: All three Standing Committee members should be voting on this. All three --

(Simultaneous speaking.)

Ms. Kyle-Lion: -- when we get to the voting for the measures that you all also, you were all full committee members on this committee. So you do

have the voting rights to vote on the measures, once we get to that point, too.

Member Jeffers: Thank you.

Co-Chair Bratzler: And Jennifer, you should of have received a link by email this morning. And when you click on that link, it'll take you to the website. We're polling everyone, polling.

Ms. Kyle-Lion: We're still currently at 13 votes. Okay, we're now at 16. And I believe we have 18 or 19 members. So we'll just give them some time.

Co-Chair Bratzler: I saw one chat that they said they couldn't make it work.

Ms. Bal: Gabby, why don't we work with Karen and Jennifer on the side to make sure they have that link and then maybe do another test before our next vote.

Ms. Kyle-Lion: Okay. We're at 18 results right now. So I can share that. Karen, were you able to make it work?

Member Glomb: And once it works, we get the message back but recorded, correct?

Ms. Kyle-Lion: Yes.

Member Glomb: Okay, got you. I wanted to double check. I see it. So I just wanted to ---

Ms. Kyle-Lion: Karen, just confirming that you were able to vote?

Okay, perfect. All rightie. Voting is now closed on the test vote. We had 94 percent say yes, they like Brussels sprouts, and six percent say no, they do not like Brussels sprouts.

All rightie, thank you, everyone for participating in that test vote. And I will turn it back over to Paula.

Ms. Farrell: Great, thank you, Gabby. All right, so next I will review the measures that we're going to be discussing during our meeting.

You can go to the next slide, please. We have one maintenance measure and two new measures that were submitted, NQF Number 3661 and 3332 were not reviewed by the Scientific Methods Panel, because they're determined to be non-complex.

The next slide, please. We do have one measure, NQF Number 3667, that was evaluated by the Scientific Methods Panel, because it was determined to be a complex measure.

Next slide, please. The measure did pass the Scientific Methods Panel review on the reliability sub-criterion but did not reach consensus on the validity sub-criterion.

So today, our Standing Committee will vote on whether to uphold the Scientific Methods Panel vote on reliability. And if the vote is greater than 60 percent in favor to uphold the vote, the Scientific Methods Panel vote is upheld.

As consensus was not reached by the Scientific Methods Panel on validity, the Standing Committee will go ahead and do a full vote on this criteria.

I also want to mention, as we get into review of our first measure, that with this fall 2021 cycle, we have created a designated timeframe for developers to respond to questions and provide clarification that the Standing Committee might ask.

So we're going to begin our discussion on each of the -- at the beginning of each measure discussion, the measure developers will have about a three to five minute window to provide introductory remarks. And then the Standing Committees will be able to discuss the measure. And any questions that come up that

are specific to the developers will be collected by NQF staff and also the co-chairs for the call.

Once the initial Standing Committee discussion on that specific criterion that we're reviewing is completed, developers will then be given an opportunity to respond to any questions that have come up during the discussion and clarify any information.

So we do ask the Standing Committee members, if you have questions for the developers as we discuss each of their criterion, please enter those questions into the chat or let us know before you verbalize your question that it is a question for the developers so that we can take note of that and ask them at the end of the Steering Committee discussion.

All right, next slide, please. And with that, we're going to get started with our review of our first measure.

Consideration of Candidate Measures

Measure 3667

Ms. Farrell: The next slide, please, NQF Number 3667, Days at Home for Patients with Complex, Chronic Conditions. And I am going to turn the meeting over to our co-chair, Dr. Dale Bratzler, to lead the discussion. Dr. Bratzler?

Co-Chair Bratzler: Yes, thank you, Paula. And thanks to NQF staff for excellent explanations.

I want to reiterate something that Paula said that's going to be a bit different than we may have done in the past. And that is as we discuss each of these sections of the consensus development process, so we're going to talk about important measurement work first.

We're going to ask that you not direct questions directly to the measure developer in your comments, that if you have specific questions you want

addressed to the measure developers, you put them in the chat.

Adam is going to be monitoring the chat for this first measure. He'll capture that information, he'll be listening to the comments so that if questions come up during our discussion, he will then summarize those. And at the end of the conversation, after our presenters have presented that section, then we will go to the measure developer and ask them to respond to the questions that Adam will help lead.

So a bit different than sometimes we've done in the past, where we've had this interactive back and forth through the discussions. We're going to try to summarize it. We will try to verbalize it so it's a part of the public record.

So our first measure is from CMS and the Yale Center for Outcomes Research and Evaluation. I believe Susannah Bernheim is here today to present. And Susannah, if you're here, we'll ask you to give a three to five-minute overview of the measure for us before we begin our discussion.

Dr. Bernheim: Hi, this is Susannah Bernheim from Yale CORE. Can you guys hear me okay?

Co-Chair Bratzler: Yes.

Dr. Bernheim: Great, okay. Thank you for your time today. My aim is to give you a brief description of the Days at Home measure, which I hope will provide some helpful clarifications about certain aspects of both the measure and the testing.

This is a measure of days spent at home for adults with complex chronic disease. Evidence shows that most patients prefer spending time at home, obviously, and in their community rather than in an acute care setting. And this measure is conceived to reflect patients' preferences.

Days spent at home are also positively associated

with better clinical outcomes and reduced cost. And there's evidence in the literature that timely delivery of appropriate primary care services, high quality care coordination, and improved care transition can all increase the number of days at home that patients spend.

For these reasons this has been a measure CoMFA (phonetic) systems supported for some time by organizations like MedPAC who did some preliminary work on a similar measure.

The measure is intended to be used in entities such as ACOs that have taken on a commitment to full-person care when they are treating their patients, because it takes that to be successful.

The intent is clearly not to eliminate all in-patient care but really to have the measure create an incentive for entities to explore or innovate ways to deliver effective, coordinated, and prevention-focused home and community-based care that produces good outcomes while reducing the patient's need for destructive and higher risk acute care.

The measure was tested in data from the Shared Savings Program ACO where it's being considered for use. But it is also being considered for use in both of CMMI's direct contracting model ACO and the Primary Care First model.

So for this reason, the application sometimes refers to ACOs when we're talking about testing data, or more generally to the provider groups that cover the types of entities that are being considered, all of which have accepted broad responsibility for cross-setting patient care.

The measured patients, the cohort, are those who have substantial disease or likelihood of high needs as defined by having CMS HVC score greater than two. The outcome is the number of days in a year that patients spend alive and at home. This is operationalized by counting the days patients are in

acute and post-acute care and then subtracting each of the days that they're alive and eligible for the measure.

Days in care are risk adjusted to account for the fact that the ACO provider entities have higher risk patients. The approach to risk adjustment modeling is similar to that used in other CMS outcomes measures using administrative claims from the prior year.

For many settings, this is a risk adjustment approach that has previously been validated using chart data. And the days at home care model, the days in care risk model also adjusts for patients dual eligibility status to account for socioeconomic factors that may impact the patient's ability to remain at home.

So that's the measure as a whole. It has one additional feature which has caused some confusion that I will briefly describe. There are concerns that a measure like this could have potential unintended consequences incentivizing patients to be at home, that is concerns about rewarding unsafe care, not having patients in the right setting. So the measure includes two additional adjustments to the overall measure, one for excess mortality and the other is for excess transition to long-term nursing care.

Just to be specific, the mortality adjustment is operationalized by using a standardized mortality ratio to estimate for each patient the excess risk of death that can be attributed to a given provider. This is adjustment to the ACO. This adjustment risk of mortality is not a stand-alone measure. Adjustment means adjusting the days in care to account for the impact of excess mortality.

This additional adjustment essentially means that, in the unusual case where an ACO or other accountable entity has substantially higher than anticipated mortality rates, this will not show up as good performance on the days at home measure.

This is a complicated issue to address and had been considered by other folks who had tried to develop a measure like this. And this approach came about through a fair amount of consideration of many different options and work with our CAP which endorsed the final formulation.

A similar adjustment is made for excess transition to long-term nursing home care. Skilled nursing care is considered an outcome of a day in care. But long-term nursing homes' time is not considered days in care. Because for some patients, the best living situation is their homes.

But to address concerns that the measure could encourage patients staying at home --- sorry, let me say that more clearly. We don't count days living in long-term nursing home as days in care. That's considered home. But we make a similar adjustment weighted less strongly in the account that there are unexpectedly high numbers of transitions for patients entering long-term nursing care. So again, I just meant to balance concerns about potential unintended consequences.

Both of these adjustments have modest to no impact on the measure scores except in cases where an ACO has substantially higher than expected rates. And as I noted, those are the result of extensive consultation with our CAP and others on how to maintain the spirit of the measure while addressing concerns about mortality and unintended consequences.

Very briefly, because I'm trying to respect the time, the measure reliability is quite high. It's based on split sample testing. The measure validity was approached in two ways. As I mentioned, it was developed with a large and very engaged CAP, 17 of 19 CAP members that responded to our face validity survey agreed or strongly agreed with face validity. Two somewhat agreed, and none disagreed.

Traditionally this has met the NQF requirement for validity. We also looked at construct validity by

comparing the measures, other measures available for ACOs that had some overlap in construct.

This is always a challenging task for outcome measures to find a perfect comparison measure that's measured in a similar population. But the measures we looked at correlated in the anticipated direction for all but one of the measures. And it was a single patient experience question that did not correlate in the anticipated direction of this measure.

Overall, this is a very feasible measure. It's based on claims data. It's intended for accountable entities that have the ability to affect many aspects of patient care. And it's intended to encourage coordinated, preventive, primary care focused on keeping patients with chronic disease at home to the extent possible.

Variation across the ACOs that we tested demonstrates the best performers have the ability to have a meaningful impact on patient stays at home. So thank you, I look forward to the conversation developing.

Co-Chair Bratzler: Thank you, Susannah. So our lead discussant for the measure is Kim Elliott, secondary discussants are Katherine Gray and William Curry. So Kim, I'll turn it over to you to talk about importance of the measure in the report.

Member Elliott: Thank you. Now, this is measure that is proposed to be for the accountable care organizations that, as the measurer developer said, there is some consideration for a provider group or provider level analysis as well in reporting.

It is a new measure, and it is an outcomes measure. And the data source is claims data which lends itself to being more reportable than some other measures that we've discussed in the past.

The numerator statement is the outcome measured for each eligible beneficiary's days spent at home and adjusted for clinical and social risk factors, risk of

death, and risk of transitioning to a long-term nursing home.

The denominator statement is the eligible beneficiaries aligned to participating provider groups and denominator exclusions they consider to be not applicable. There are currently no exclusions or exceptions for the measures. So all patients named in the denominator inclusion criteria are included in the reporting.

The importance of the measure, we do need measures that --- we don't have a lot in this category. So measures that keep people at home and report the ability to keep people at home based on care coordination is something that is really considered valuable at this time.

So I'd like to open it up now to the -- to see if they have anything that they would like to say in addition to that for the introduction.

Member Curry: I have nothing to add. Thanks.

Member Elliott: Katherine, anything?

Member Gray: I have nothing to add.

Member Elliott: Okay. As far as an opportunity to improve, the panel did say that it was moderate. And I agree with that. There are, based on the data that was presented, it looks like there is an opportunity to improve.

And quite a few of the different accountable care organizations that were included in the study, and then there were others, of course, that performed at a little bit higher rate, so there is an opportunity for improvement.

Co-Chair Bratzler: So I think, I want to make sure NQF staff keep me -- because that's actually the first thing is voting on the evidence, and the second thing is the performance gap. So evidence there, I didn't hear a lot of conversation about.

Member Elliott: Yes.

Co-Chair Bratzler: I know that in the initial conversation, the description of the measure, there is discussion of the evidence that patients who are not -- there may be linkages between this particular performance measure, and patient outcomes, and other things. And I get it with respect to patient experience, but was there strong evidence? What is the strength of evidence is the question I'm asking.

Member Elliott: Are you moving into the evidence, Dale?

Co-Chair Bratzler: Yes.

Member Elliott: Okay. Do you want me to go ahead with that?

Co-Chair Bratzler: Yes, I think so.

Member Elliott: Okay. According to the information provided by the measure developer, they did use a logic model that indicated that timely access to high quality preventive and primary care, and consideration of patient preferences for care settings, and improved care coordination, care transitions, led to more patient time spent at home and reduced over utilization of acute and long-term institutional care settings.

The developer also did a pretty comprehensive literature review of relevant peer review publications and found that most patients and their families prefer spending more time at home and in the community, which is what the days at home is in this particular measure, than in a hospital setting or in an institutional setting.

Poor care coordination can lead to unnecessary and preventable hospital visits for patients. In contrast, improved care coordination and care transitions prevent unplanned hospital visits leading to more days at home and higher quality, timely care.

And the literature also, according to the developer, indicated that, given that patients' complex chronic conditions often receive care from several clinicians and sites of care, this patient population may particularly benefit from improved care coordination.

And then did you want me to go on to the opportunity for improvement, Dale?

Co-Chair Bratzler: No, let's go ahead and have the discussion about evidence and see if Katherine or Dr. Gray have any comments about evidence. And we'll take those votes individually.

Member Elliott: Okay.

Member Curry: So Bill Curry. So the literature review was limited to the last ten years which I think is a real strength of the literature review.

But when I looked at the studies that were cited, there was one call to get a systematic review in advanced cancer. There was one review about hospital at home. There was one about patient-centered care assessments regarding claims, three in palliative care and end of life, four in care coordination and integrated care, and six in stroke and intracranial hemorrhage.

So I think that, you know, they have a nice list of evidence-based reviews that they've looked at. But I just myself question was the scope of disease large enough?

We talk about advanced cancer, we talk about palliative care, end of life, we talk about stroke and intracranial hemorrhage. But where is the information on chronic kidney disease, advanced diabetes complications, other cardiovascular disease? So that was a question that I had in my mind as I looked at the evidence. But they provided some nice information with what was given.

Co-Chair Bratzler: Thanks. And I think that's an

important one that we'll keep in our -- that Adam is keeping track of here before, when we get to that portion.

Katherine, did you have any other comments about the actual evidence itself on the metric?

Member Gray: I don't know if this is the right place for it, but I was curious. There's, you know, the question sort of at the end of passing the evidence is whether the providers can have, you know, one action that they can do to improve the care.

And the Scientific Methods Panel said yes in the past. I was just curious. And maybe others, you know, have a more clear understanding of how the various providers can impact some of these decisions.

And therefore, I didn't feel like it was really clear to me that there was evidence that there were actions that could be taken, you know, to make some kind of difference in whether somebody had more days of care versus less, depending on perhaps, you know, if they have an acute episode, then they get admitted to the hospital.

I mean, I'm just trying to figure out where in there is the practical difference of what providers can do or, and that sort of brought me to another thing, my confusion over the ACO as the measurement unit or the outcome unit versus the providers. And in particular, I went back to confusion about if there were 610 ACOs, how many provider groups does that represent? Or are they synonymous with each other?

Co-Chair Bratzler: All right, thank you, Katherine. Jeff Susman, you had a question?

Member Susman: Yes. It continues on with the previous conversation. I can see it in ACO level, how this would be a good global measure, perhaps. I have a lot harder time though buying that this is a credible evidence linked measure for primary care groups, for contracting, or providers, or at the provider group

level.

I think that the evidence that has been cited is largely downstream and indirect. And it also, while attempt has been to adjust for unintended consequences for keeping people at home that would otherwise be better served in hospital settings or other acute settings, we have to buy a black box for that. And I'm not a big black box fan.

It's an extremely complicated issue to adjust for differences in baseline populations. And I honestly am very skeptical that every ACO population is similar and can be adjusted appropriately to accurately reflect the quality of care. So those are my comments. Thank you.

Co-Chair Bratzler: Thank you. Anna, I see your comment, and we'll keep that in the queue for the measure developer.

Any other questions that somebody hasn't put into the chat for the measure developer? Then I'll turn it over to Adam to guide that part of the conversation.

Member Glomb: Dale, this is Brendle Glomb. I just wanted to just comment as well. I just want to make sure we will be getting into, as directed at the beginning, the validity and the --

Co-Chair Bratzler: Absolutely.

Member Glomb: -- liability. Right, okay. Because I have some big concerns that are specific and fall in that area rather than voicing them now.

Co-Chair Bratzler: Yes. Absolutely, that --

(Simultaneous speaking.)

Co-Chair Bratzler: We'll get to that for sure.

Member Glomb: Thank you so much.

Co-Chair Bratzler: Right now we're talking only about

the importance of the measure.

I think David Lang, are you on mute, David?

Member Lang: Yes, sorry. Well, similar to what was just said, you know, the measure is intended to achieve a worthy goal, but there are a lot of moving parts here, and I guess this is it, also in the realm of validity.

But as long as it's been mentioned, I'm concerned there are no solutions for the measure. It seems to me that in some cases patients would be better off in the hospital rather than elsewhere. And days at home conceivably may encourage earlier discharges when these may not be appropriate, including increased risk for readmission and lead to untoward outcomes.

Co-Chair Bratzler: Okay, thank you. Adam, I think I'm going to turn it over to you to lead the questions for the measure developer to respond to.

Co-Chair Thompson: Great, thanks so much.

So I've got four questions here. So I'll begin with the first question. It came from Harold Pincus. Does the measure take into account the local supply of skilled nursing facilities and hospital beds?

Dr. Bernheim: Right. Can you hear me?

Co-Chair Thompson: Yes.

(Simultaneous speaking.)

Dr. Bernheim: It did not get incorporated into the measure. But concern there might be that, in areas where there's a lot of hospital beds, we physicians have a tendency to just be more likely to admit. And obviously that's not an ideal situation.

And we didn't look at, just between local supply and how providers do on the measure, there's not a relationship with hospital beds. There is a slight relationship with nursing home beds.

And again, the concern there is that we, you know, you wouldn't want to adjust for the fact that there are more nursing home beds as if that alone was a reason for an ACO to have higher days in using of those beds. So it's not included in the measure.

Co-Chair Thompson: Thank you much. Our second question is, picking up on what Bill was talking about, what about the other types of chronic conditions that weren't specified, such as diabetes, chronic kidney disease, et cetera?

Dr. Bernheim: Great question, right. But it's actually just a couple of things. One is the place where we think we can have the most impact is on patients who have more chronic illness, because they're just more likely to need that acute care.

And we're more likely to therefore, through best practices, keep them healthy, keep them at home, make it less likely that they're admitted to the hospital and more likely that we address things in an outpatient setting.

So the measure, again, is focused on patients with an HCC score of greater than two, that's to align with how there are some other patient definitions within those CMMI models that use that same scoring.

But just to give you a sense, to get a score of greater than two, I think that's among the sickest of the Medicare patients, not the top two percent but sort of in the order of 20 percent. So they have a wide variety of chronic conditions.

And the concept is, although as a primary care physician I can't prevent every hospitalization, mostly what I'm doing most days is trying to contribute to decreasing the risk that my patients are going to end up in the hospital. And I can't get it to zero, and I certainly can't do it alone.

So that's why it's intended for a broader population, broader care providers like ACOs and larger groups

that can work together to lower the risk, to bring down the risk that patients are going to spend a lot of days in care, regardless of the mix of their diseases. And we can try to bring forth literature and other fields if that would be helpful.

Co-Chair Thompson: Thank you much. The next question came from Katherine which picks up a little bit on what I think you were just defining.

Are ACO and provider groups synonymous?

Dr. Bernheim: And this is where we were trying to be careful, and we have a lot of confusion.

The measure was developed in data from the Shared Savings Program. The DC model, as folks know, is essentially a next generation, not next generation but a next generation of ACOs.

The Primary Care First model, which may use this measure is a little bit different. But again, in the context of a voluntary group, a voluntary choice to take on broader responsibility for patients, it is not intended for an individual primary care physician. It's intended for the context in which the --- and we used the word provider group to mean ACO, but PCS is not officially an ACO. So we were sort of trying to hedge our bets.

But the intent is that it be used in these contexts where there are a group of providers, where they are not officially an ACO, who have elected to take on cross-setting care of a large population of patients. I hope that helps. I'm trying to get the words right to sort of reassure folks about this.

Co-Chair Thompson: Great, thank you so much for the clarification.

Next we have two more questions, I think, one more on my list and one developing in the chat room. Is there evidence that this measure is actually a valid surrogate for care coordination, from Anna.

Dr. Bernheim: It's a great question. And I'll tell you, directly measuring care coordination is a challenge, right. So we sometimes have to get at these things indirectly.

But I'll give you two quick answers. One is we try to compare it with measures that are thought to be important signals of care coordination like admission specifically for things that are ambulatory incidents that we see a relationship with this measure and avoidance of ambulatory-sensitive conditions which is a much narrower construct. So it doesn't affect patients as broadly.

But it also, you know, has some history behind it of being really focused on the things that we can do. And yes, I mean, the measure is based on this concept that, you know, what we accomplish when we coordinate care, even if we can't precisely measure every time, you know, I follow-up with a neurologist or the, you know, the test results of where it's supposed to be at the right time.

It's that it's a sum at the end of day, when we do these things right, what our patients experience in fewer times that they get to the point where we have to admit them to hospital as opposed to being able to manage things in coordinated fashion.

So we did our best with the measure. They're available to show that it correlates with that, and we lean on the evidence that, when we do these things well, we do see that patients spend less time in the acute care setting.

Co-Chair Thompson: Thank you so much. And then did you look at geography, morality, urban frontier, from Jeff Susman.

Dr. Bernheim: We thought about this a lot. We did. I am trying to remember. So can we jump to the next question and then come back to this, because my team is going to tell me in a second what I have about morality. But I don't have it at my fingertips.

Co-Chair Thompson: Absolutely. And let me just jump into the chat room here one more time, see what we've got here. I see a question from Jennifer. How is palliative care hospice provided in a skilled nursing --- oh, how is palliative care in hospice providers in a skilled nursing facility considered, from Jennifer.

Dr. Bernheim: Great question. So we, our techs felt strongly that once a patient enters hospice, everything going forward, regardless of where the care is provided, is considered at home. But that should not --- the days spent, whether you're in an in-patient hospice or if you're elsewhere, so once you're on hospice, that's considered being at home. And it's a good outcome so as not to discourage that.

And just to say we looked at urban versus rural and did not see a relationship with days in care. Obviously, there's different challenges for different providers, but it doesn't seem that the measure particularly disadvantages either setting, given the risk adjustment approach.

Co-Chair Thompson: Thanks so much. So I want to do a quick check in here on our question process. Some of the questions we are getting coming in are information that's contained somewhat in the measure.

So what we want to do is, as those questions come up, we're going to steer them back to the committee to answer first to see if we have the information. And then if our lead discussant, or supporting discussant, or other committee members don't have a response to that, then we'll turn back to the measure developer. Sound good to folks? Awesome.

Okay, so I believe, let me check in the chat room. There are two more questions that have come up, so I'll kick this first to the committee. How does the risk adjustment correspond to RA done under CMS system for MCR patients?

Dr. Bernheim: I know the acronym RA, risk adjustment, but I'm afraid I'm not sure what MCR is. Can someone just tell me what they're asking?

Member Glomb: It's Medicare, I'm sorry, it's -- yeah, I was just trying to be brief. How does your risk adjustment, particularly with the risk of death, other things that are added in here, correspond with the overall CMS risk adjustment measures that are done on all Medicare patients on a yearly basis?

Dr. Bernheim: Right. So for other outcome quality measures the CMS uses the fundamental risk adjustment is pretty similarly in that we utilize comprehensive claims from the prior year so that we're seeing diagnoses that may've just popped up in an outpatient setting or in the hospital setting.

We group them according to a commonly used grouper and we do a fair amount of testing to decide which ones to include.

In this case we actually modeled it after a similar measure where a lot of that testing and the risk adjustment approach had been used.

So in that way it's very similar and, as I mentioned in the intro, there has been some claims-based validation of that risk adjustment approach.

The thing that is unique about this measure is the additional adjustment for excessive mortality or nursing home transition in a given ACO. That's unique to this measure.

Co-Chair Thompson: Thanks so much. I believe we have one more question in the chat room from Harold, and I'll pose this to the lead discussants and supporting discussants first.

What about availability of family and other unpaid care givers at home, was that in the measure anywhere, folks?

Member Elliott: I did not see that in the details

provided.

Co-Chair Thompson: Okay. So now I'll kick it over to the measure developer.

Dr. Bernheim: Yes, just very briefly, you know, this is data that would be tremendously useful and is not routinely available for Medicare patients, so we could not address this.

The part of the underlying -- so there's two pieces about that, one is, we tried not to, in general, penalize the transition to long-term nursing care, because sometimes that represents a transition that has to do with having fewer social support resources. But if you transition to long-term nursing care, that do not start counting as days in care because it's considered a transition, and then the adjustment allows us to ensure that's not happening at an excessive rate in a given ACO.

But anyway, that was, the fact that was didn't address that was part a, the decision to not consider days when you transition to long-term nursing care as all days in care. Because that could penalize situations where there's just not the unpaid social support.

The other thing is that part of our risk adjustment for dual eligibility, which is imperfect, but an important marker of -- especially if you're in care and so less likelihood to be able to sort of bring in your own private care to help you stay at home.

So those two things were meant to address that concern, because we don't have perfect data on what kind of supports people have in their house.

Co-Chair Thompson: Great. Thank you so much.

So I think what we're going to do now is give our measure developer a chance to catch their breaths over there and I'm going to pass it back to Dale.

Co-Chair BratzlerBRATZLER: Okay, Adam, thank you.

So I want to see if any of the committee members have any other comments before we take our vote on importance to measure.

So the evidence, focusing on the evidence.

All right. Let's open up the vote on evidence.

Ms. Kyle-Lion: Give me one second. Okay, voting is now open for measure 3667 on evidence, the options are, A, for pass, or B, for do not pass.

Ms. Bal: Gabby, I think we had a few people join the call so I just want to make sure that we have everyone's disclosure before we finish with the vote and --

Ms. Kyle-Lion: Sure.

Ms. Bal: And I see we have a couple left coming. Has any abstaining committee joined since our initial introduction that need to do an introduction and disclose anything about their work?

I think we saw two committee members join.

Heidi, are you on the phone, or Jette?

Okay, never mind, I guess 18's still our number, thank you.

Ms. Kyle-Lion: Okay, we have 18 votes in.

We have 14 votes for pass and four votes for do not pass. Just give us a moment to calculate.

Therefore the measure passes on evidence. Thank you.

Co-Chair Bratzler: All right, so we'll move on. Kim, if you'd like to talk about performance gap.

Member Elliott: Certainly. So the evidence or documentation submitted by the measure developer referenced back to 2017 to 2018, Medicare fee-for-service data that included 610 ACOs.

And the average adjusted days at home were 330.4 days and the range was 291 to 345.9.

The developer also referenced different studies that demonstrated substantial variation in time spent at home and suggested that there is an opportunity to improve the quality of care, and the resulting days at home for the target population.

As far as disparities, there was some evidence presented by the developer of disparities related to age, Medicare, dual eligible status of Medicare/Medicaid members.

Minor disparity was shown in the evidence for social risk, inconsistent study results were also shown in the evidence related to a link between the disparities and the socioeconomic status.

And then there was some discussion in the evidence submitted about proper location, that some people might be better off in a skilled nursing facility, a nursing home, even a hospital, rather than being at home.

And we had some feedback that indicated that the disparity data that was presented by the measure developer was not real clean, it was a little bit messy.

Did you want me to talk about the pre-evaluation comments also at this time?

Okay, the pre-evaluations comments, two community members expressed concern that the outcome may not be related to what the patient refers to as outcome, so non-facility based care.

The evidence also showed that there was a large variation in the days at home across ACOs, suggesting that the organizational level, there are interventions or processes that can be fortified to keep people at home.

Another member commented that there appears to be a moderate amount of medical evidence to

suggest that this process measure is associated with improved quality.

The measure also indicates that appropriate care transition such as, case management, preventive and routine care, and follow up care will result in more days at home versus in a long-term care setting.

In most cases this would be an appropriate relationship of actions in care of service and outcomes being measured.

And did you want me to talk about the performance gap also?

Co-Chair Bratzler: Yes, that's what we'll be voting on, is performance gap.

Member Elliott: Okay. So it would appear that the gap in care in the days at a facility versus days at home is more provider group variable and risk adjustment variable than representative of a global sub-optimal performance standard.

And that was a comment that was echoed by quite a few different people that were commenting on the performance measure and the performance gap.

The measure developer provides in evidence that the gap exists for days at home ranging from the 291 to 346 for the standard deviation of 3.7 days.

Developers presented data demonstrating variability and do provide in some evidence of disparities, but the disparity evidence was a lot weaker than the other evidence.

The performance gap analysis utilized Medicare ACO data only and yet the measure, when you look at a lot of the information that was written, it was considered for all Medicare patients, not just the MCOs or the ACOs.

So that's something that we really need to think about from a gap in care perspective and whether it

really is applicable to both or just one of those populations.

And although the measure has a risk adjustment, and they actually included three different risk adjustment models, a performance gap may not be able to be identified.

And the study results had conflicting results regarding disparity impacts or disparities in care.

The average range of days at home is substantial but it is not clear if the variability among the ACOs provides a similar performance gap.

And the patients are evaluated individually for the risk adjustment and some of the comments and feedback from our people that reviewed it was, not sure how that could be managed solely by the accountable care organization, and what other entities might be involved in that.

And I think I'm going to stop there and see if our co-discussants have anything to add.

Co-Chair Bratzler: Yeah, so Katherine and Bill?

Member Glomb: I find it interesting that the dual eligible population has some differences, and perhaps that's related to disparities of social determinants, socioeconomic status, and yet in the indicators for social risk they were found to be -- the ones that they looked at -- were not statistically significant, or only of modest impact.

So my challenge is that the disparity information is not convincing and I think there's a lot of work that needs to be done in the future.

Just last week, NIH came out with a R01 actually to look at this very topic, so in the current model I have some challenges with how the disparities play in.

That's what I would have to mention about this, I think what we're trying to measure is important, how

do the disparities play across various ACOs and provider groups.

Co-Chair Bratzler: All right, thank you. Katherine, do you have any comments?

Member Gray: Yeah, this has lots and lots of comments, as Kim has brought forth, but I had questions about other things that might be of interest.

Like, how much of the risk adjustment is based on -- or was it looked at -- for the size of the provider groups, you know, either ACO or provider groups direct contracting, or however it all fits together.

And I guess the answer is that the 610 is the same for provider groups as it is for ACOs, or is that still confusing, if it's really technically an ACO are there more provider groups in there?

But, anyway, the size of the provider group could also be a factor.

And then also I wondered, had the developer looked at the risk for readmission in the, you know, regular Medicare compared to what they have in, you know, in their pool inside the ACOs.

Co-Chair Bratzler: So, just one clarification here, and thank you for reminding me, that we will discuss risk adjustment substantially when we get to particularly reliability and validity. So we're going to talk about that.

So our principal conversation and our vote will be on performance gap.

And I know all that goes together, I get it, but that's what we're going to be looking at first is performance gap.

So I'm going to see if any members of the committee have questions for our discussants and then certainly any that you want to put in the chat, so Adam can

leave that with our measure developers.

But --

Ms. Bal: And Dale -- sorry, Dale, if I could just provide a little clarity, I know that disparities -- the disparity section and the risk adjustment section often get confusing about where to discuss what.

When we say disparities under performance gap, we're referring to, is there a gap?

So let's say we're looking at performance gap and everything looks great, everybody's performing really well, but then when you get into disparities you start to see, oh, there actually is variation in the performance when you look at disparities.

And so that's why we include that in performance gap, so we can make sure that care is universally being provided at the level that we want.

And so that's what you're looking for in disparities, and then when we get to risk adjustment, that's when you think about, you know, what's included, what shouldn't be included and, you know, how do you actually account for those disparities.

So hopefully that provides a little clarity about the differences there.

Co-Chair Bratzler: Yeah, thank you very much.

So any of the committee that have questions for our discussants, and then I will turn it over to Adam to?

So I know a number of questions or things have been posted in the chat so, Adam, I'll turn it over to you.

Co-Chair Thompson: Great, thanks so much.

So just real quick to review the process what I'll do is, ask the question, take it first to our lead and supporting discussants, and then if we also have our hands in the air and question marks on our face then

we'll turn it over to the measure developer.

So the first question that came in, and this is a general question for the committee from Anna about, if anyone is aware of what percentage of the Medicare population is represented in the ACO model?

I see hands in the air, there we go. Measure developer, do you have any data related to the number of Medicare folks that are in the ACO model?

Dr. Bernheim: No, I was really hoping your committee was going to answer that question.

I don't and we can tell you how many -- in the form you can see how many patients were included in the testing, so that gives you a sense of how many were in our group that was tested.

But I don't know it as a proportion of the Medicare population and, again, remembering that we were testing in the Shared Savings population, but it's being considered for use in other groups as well -- in other ACOs as well.

Co-Chair Thompson: All right, thanks so much.

So Anna, you and I will have to Google soon -- oh, well, Bill has his hand up. Maybe --

Member Curry: So, yes, Google is awesome. It says -- from a group that says that 20 percent of all Medicare patients are in ACOs and only a third of the traditional Medicare patients are in ACOs.

Co-Chair Thompson: Awesome, thanks. So we will not count that as the most rock-solid evidence that we've ever had in the whole, wide world, but I think it does give us a general scope to move forward. Thanks so much.

Our next question, again, kicking to the lead and supporting discussants first, how much of the gap that was presented was related to social determinants of health?

Member Elliott: That was the interesting thing in the documentation that was submitted, it didn't appear that were was a real strong link related to social determinants of health.

Co-Chair Thompson: Thanks so much, Kim.

Co-Chair Bratzler: Yeah, so Kim, I actually noticed that too, when I read through that.

I'm not sure, though -- I guess -- this is my personal experience rounding is that, without a social worker at the bedside, I'm not sure we can risk adjust for social determinants based on things that are easily captured from claims or other things.

And more things like area deprivation index and other things used in those models.

Member Susman: I mean, my question, just to clarify is, one ACO serving a population with a high burden of social determinants of health that are negative, is it fair to compare the outcomes, days at home, with, say, a suburban well-serviced with lower social determinants of health?

And maybe for the developer to respond to whether social determinants of health were somehow adjusted for when we get to that point. Thank you.

Co-Chair Thompson: Thanks so much. I see -- Bill, you have you hand up?

Member Curry: So my question would be, how many groups are actually measuring social determinants using a standardized approach, such as the PRAPARE tool.

And if they are, how is that information put into the EMR for each individual and then be able to collect it at the ACO level.

Versus, you know, looking at census track data or other, you know, geospatial issues around social determinants.

So I think it's a challenge with being able to get this data from the EMR and brought up to the ACO level.

That's my question and concern.

Co-Chair Thompson: Thanks so much.

And I believe our last question -- we had another about risk that we're going to hold until we get to that section.

And then I believe, Jeff, you had another question about the average days at home seems to be skewed upwards, is there good evidence that the lowest deciles better care, should everyone be striving two-hundred something?

So I'll kick that first to the lead and supporting discussants.

Member Elliott: I actually had a very similar concern as I was reading through the evidence that was submitted.

The application of how they're defining at-home seems different than what I've seen in other measures, other studies.

So I think there probably is some potential for skewing, but I'd like to open that up to the other discussants.

Co-Chair Thompson: Thanks, Kim.

Any of our other supporting discussants that would like to share their view on this?

Member Gray: This is Katherine. Yes, I agree, especially the sort of residential nursing home situation being counted as days at home.

I have to ask how much control that there is over those people that are in, like, intermediate care or whatever. Just seems unusual.

Co-Chair Thompson: Thanks.

I believe -- correct me if I'm wrong, Dale, but I think that's our last question in the chat room.

I think you're on mute.

Member McCollister: Adam, can I ask a question? The thing that's kind of, like, bothering me about this, my father died in late 2019 -- really complex chronic illness, in and out of the hospital.

The days that he was not in the hospital is because my mother was doing all the coordination, he was not in an ACO.

But the thing that, again, speaking from the patient/family member perspective here, as opposed to a physician, I feel like there's a lot of work to be done by people who are not coordinating, like, the individual care givers or patient to keep themselves or their loved one out of the hospital.

And I'm just wondering if we're giving doctor's credit for unpaid work that the family is doing.

And that's what's -- I mean, I don't know if this is a reliable measure of the care coordination for physicians as opposed to the care coordination that all of these unpaid workers and family members are doing.

Co-Chair Thompson: Thanks so much, Anna. That's a really great point. I just was coordinating with my husband and his family as his grandfather was dying.

And it was a whole family healthcare center that we opened to make that possible.

Member McCollister: Yeah, and I spend a lot of time coordinating my care to stay out of the hospital, I have 13 physicians, none of my doctor's do it.

I'm well enough to be able to do it, I'm not on Medicare, but it's a real issue and I would like to find quality measures that address that concern and burden.

I mean, anyway, I know that's not the measure that we're looking at, but I'm just wondering if we're actually measuring the thing that we want to measure.

Co-Chair Thompson: Great, thanks so much.

So Dale, I think that's all our questions. I can kick it back to you if you want to facilitate it.

Co-Chair Bratzler: Yeah, so I think so and unless any of the committee members have any other questions related to gap, we'll move on to the vote on gap performance.

Co-Chair Bratzler: Raquel had a question, and I think the developer also asked to respond, but let's go Raquel first.

Co-Chair Bratzler: Okay.

Member Jeffers: Thank you. So to plus one to what Anna said, I think that is more the rule than the exception.

So really would love further clarification on whether the measure is just applied to members who are in an ACO relationship where we know that at a minimum they're getting some kind of care coordination, or whether the measure will be applied to the entire Medicare population.

Co-Chair Bratzler: Well, so Susannah can respond, but I will say, in general when we evaluate measures we're not telling the developer, which in this case is in part same as how they use it.

Although I think the measure was developed and submitted as an ACO measure and I, like you, feel that extending it to other models like primary care first, which is a whole bunch of private practice docs that signed up.

That, although there's an expectation for care coordination, they don't have the resources of an

ACO, I have strong problems with.

But I'm not sure how much of that's within our purview here. We're to look at the strength of the measure.

Dr. Bernheim: So I'll just briefly review the role that family care givers play in coordination.

You know, and in agreement that there's a lot of questions that I would love us to be able to ask and measure about the burden on family care givers.

But without in the measure that holds the ACO accountable for trying to put this in the home, we are essentially saying that it doesn't have responsibility. The pull for a measure like this is that now the ACO feels some responsibility for helping, and it takes some of that responsibility -- it holds some of that accountability into the provider that you're saying is missing.

It's not perfect and I agree that there's really an unseen burden that this measure cannot perfectly ameliorate.

But just to be clear about the intent is to say, yes, this is part of the responsibility of a accountable entity that has taken on coordination responsibilities and if they're doing their job well it should provide more services to help with this issue.

I don't want to put CMS on the spot, and I don't actually know who's on the call, but there's really important questions about the intended use in other models.

And so we'll pause to just see if any of our CMS colleagues want -- well let me first ask, Dale, would you want to hear from CMS or, as you said, it's generally not --

Co-Chair Bratzler: It's really not our purview to decide how the measure gets used, it's -- whether or not the measure --

Dr. Bernheim: Right.

Co-Chair Bratzler: I mean, it's important because I feel very differently about this measure when I think about our primary care, you know, PCP providers versus doing it as an ACO where there's a lot more analytical resources and everything, so.

But, again, I don't want to cut that off but I just point out that we're to talk about the measure -- consensus on the measure.

Member McCollister: I do think it's relevant to the applicability of the measure.

I mean, if they're going to use this outside of the setting for which it was designed, and from which the data was collected, that's different.

Co-Chair Bratzler: And, Anna, I completely agree with you, I think we'll get to more of that conversation subsequently. Right now we're talking about, is there a performance gap.

That's what we need to vote on next and so I'd like to keep our comments there.

By the way, Adam and I both have talked about, we thought this measure was the one that was going to generate the most conversation today.

Are there any other comments by committee members particularly related to performance gap?

We've heard what was provided by the measure developer in the measure submission, so is there anything else on performance gap that we need to discuss?

Okay, hearing none I guess we will ask that NQF staff open that vote.

Ms. Kyle-Lion: Prior to opening the vote, we did notice that committee member Jette Hogenmiller joined and was commenting in the chat so we just

ask, Jette, if you could unmute and provide any disclosures that you have at this time?

Jette, if you're speaking, we can't hear you, so maybe message in the chat any of your disclosures

Okay, we're going to move on to voting. Just give me one second to share my screen.

Okay, voting is now open on measure 3667 for performance gap, the options are, A, for high, B, for moderate, C, for low, or D, for insufficient.

I'm seeing 17 votes, I believe we're waiting on one more.

Okay, I see we've reached 18 votes.

Voting is now closed on measure 3667 for performance gap, we have one vote for high, 10 votes for moderate, two votes for low, and five votes for insufficient.

Therefore, the measure passes on performance gap. Thank you so much.

Co-Chair Bratzler: All right. So Kim will go ahead and talk about scientific acceptability, particularly about reliability.

And here we're going to make a decision about whether we accept the scientific methods panel for reliability.

We need to vote on that first, I think.

Member Elliott: I can provide a little bit of a summary of what the scientific methods panel found.

Co-Chair Bratzler: Yes.

Member Elliott: And in the preliminary analysis the sub-group members found the specifications confusing and occasionally arbitrary with little evidence for the measure constructs.

There was potential misalignment of concept presentations within the submission, and noted the denominator statement lacked an explanation of the target population, conditions, settings, and other pertinent measure constructs information.

They were concerned that several concepts included in the submission were not documented as exclusions in the specifications which both, threatens the measure's validity and may incentivize under treatment of conditions, potentially outside the locus of control of the accountable entity.

Including very low outliers that can never reach the expected performance gain for permanent nursing home residence, etc.

Panel also questioned whether the consideration of exclusions included patients treated in emergency departments admitted to acute care settings and days after death occurs would always indicate a low quality care.

The panel also expressed concerns with adjusting for transitions to the nursing home, which purports that moving from home to nursing home is always negative.

And other concerning data elements included permanent nursing home admissions requiring skilled nursing care, which may include personal and community resources that are not modifiable by the accountable care entity.

The panel also noted that the unit of analysis reported in the measure vacillates between accountable care organizations and provider groups.

And one panel member questioned whether this measure, which combines multiple risk model calculations into a single overall score, should be consider a cost-to-composite measure.

Co-Chair Bratzler: I just want to clarify, so I'm

looking through the submission forms, and the scientific methods panel gave moderate on reliability and then did not come to consensus on validity, as I recall.

Member Elliott: That's correct.

Co-Chair Bratzler: So just a point of clarification for staff, should we vote on just whether or not we accept scientific methods panel for reliability and then move into the conversation about validity, which we have to have?

Ms. Bal: The developer did note that there had been some additional information provided to the SMP after that feedback.

I wanted to just elaborate on that a little bit before we make that vote of accepting SMP or not.

Just to make sure all the current information is available, would that be okay, Dale?

Co-Chair Bratzler: Yeah, that's fine, I just want to clarify though, we're really only voting on their acceptance of reliability, not -- the validity, I think we have to have the full discussion.

Ms. Bal: Exactly.

Co-Chair Bratzler: Okay, yeah, so Susannah, if there are other comments you want to make, particularly about reliability.

Dr. Bernheim: Yeah, just two very brief comments. And so one is the concerns from the Scientific Methods Committee that were just right off have been updated and corrected, so I'm concerned that this committee may not have the full information.

Our understanding was that the NQF staff updating around sort of the extent of these concerns was going to be provided to you, and I think some of it is addressed in the later public comments from our team. So I hope that this committee will have the

ability to sort of understand a little bit better how these things got parsed out as part of your discussion.

But to the point about reliability, the Scientific Methods Committee was supportive of that, so we can -- happy to provide clarifications about how we do our reliability assessments if people would like. But I think the other pieces of information probably come up more in validity, and I just want to make sure that you all are getting the full representation of the concerns.

Co-Chair Bratzler: Adam, I've seen a couple of comments about reliability.

Co-Chair Thompson: Yeah. I've got two questions here. One is, was reliability tested in non-ACO setting? So I'll kick that first to our lead discussant.

Member Elliott: The documentation provided did not indicate that it was tested outside of the ACO setting.

Co-Chair Thompson: Great. Thank you, Kim. Measure developer, would you agree with that?

Dr. Bernheim: Yes. So, you know, as tends to occur, a measure developed and tested in a group that is meant to be representative of the types of groups that would use the measure, but we don't necessarily test it in every population that it will be used in. So all of the measure development and testing that you have was on the 610.

Co-Chair Thompson: Thanks so much. The second question, which stemmed from that, which was could they discuss how reliability was tested if it was tested beyond ACO?

Co-Chair Bratzler: Which they said it wasn't.

Co-Chair Thompson: Just -- no. Yeah.

Co-Chair Bratzler: Yeah.

Participant: Just got it. Thank you.

Co-Chair Thompson: Perfect. All right, Dale, back to you.

Co-Chair Bratzler: Yeah. So I think at this point, the best thing for us to do is to go ahead and have the vote on whether we're going to accept the Scientific Methods Panel's evaluation on reliability, and then we can move into the validity conversation.

Ms. Kyle-Lion: All righty. Give me one moment to share my screen. Okay. Voting is now open for Measure 3667 on whether you all as the committee accept the Scientific Method Panel's rating for reliability. The options are yes or no.

Member Jeffers: And just to clarify, the Scientific Panel voted and approved the reliability and -- endorsed the reliability.

Co-Chair Bratzler: Reliability they gave it a moderate rating.

Ms. Kyle-Lion: Okay. We got 18 votes. Nineteen votes.

Ms. Bal: Nineteenth vote is Jette. I know we've asked for your introduction disclosure a few times. Jette, if you can speak up, that would be great. If you can't, please put your introduction disclosures in the chat. Based on what we have, you don't have any. But we would just like to make sure that it's verbally disclosed on the call.

Co-Chair Bratzler: All right, she has responded in the chat, so we can get the summary.

Ms. Kyle-Lion: Okay. Perfect. Okay. Okay. So 74 percent have voted yes, and 26 percent have voted no.

Co-Chair Bratzler: Okay, thank you. So now -- I'm sorry.

Ms. Kyle-Lion: Sorry. No I was just going to say, therefore, you all accepted the Scientific Method.

Co-Chair Bratzler: The reliability, yes, so --

Ms. Kyle-Lion: Yes.

Co-Chair Bratzler: -- now we have to have the conversation about validity where the Scientific Methods Panel did not come to consensus. This is a must pass criteria, and so, Kim, I'll turn it over to you.

Member Elliott: Thank you, Dale. So the developer conducted construct validity with Pearson correlations to six other ACO level measures hypothesizing that the quality conceptually relates to excess days in care for patients with complex chronic conditions.

According to the information provided, the Pearson's correlations did not correlate well, and they ranged from a negative 0.549 to a positive 0.048 resulting in a high inverse correlation for unplanned admissions, moderate correlation with other measures, and no correlation with fall risk, and an unexpected inversion correlation with patient experience.

The developer did explain that it was possibility due to hospital admissions and readmission measures. And the developer also reported the poor correlations may result from testing against measures using smaller sample sizes in which we're not risk adjusted for the clinical variables.

The developer also performed face validity testing of the computed measure score, and consisted of 19 of 21 responding members who assessed whether the days-at-home measure as specified can be used to distinguish between better or worse performance at ACOs or provider groups. Two members indicated strongly agree, 15 indicated agree, and two indicated somewhat agree.

As far as the risk adjustment, some of the Scientific Method Panel members know that there are three different risk adjustment models used. They expressed concern about lack of clarity and about whether or how they were combined to get a single score and the validity of the approach.

They also had some concerns with the model construction which they agreed lacked vital adjustment and consideration for many variables without theoretical and empirical justifications, and used arbitrary measure waiting.

Specifically the unexplained selection of waiting mortality days at 1.25 percent, and the annual nursing home start date of January 1st. But are not conceptually or empirically demonstrated or justified. And the developers did acknowledge that these were not empirically assessed, but rather are subjective and based solely on recommendations.

A few of the Scientific Method Panel members discussed the effect of specific chronic conditions on the risk model such as cancer, dementia, and congestive heart failure that increase either by nature of these disease states.

And the greatest concern for the risk adjustment model expressed from the Scientific Methods Panel members was not the development approach for days at home -- or was the development approach for the days at home and the mortality and nursing models.

The exclusions were also discussed by the Scientific Methods Panel, and they questioned the process outcome pathway that resulted in increased rather than decreased days of care, and the lack of exclusions for long-term nursing home residents prior to a measuring period who have no chance of at-home days defined in the specifications.

The Scientific Methods Panel also indicated the discrimination and calibration were generally

acceptable, but had concerns related to low outliers. And they attributed that or connected that to an unintended consequence of the measure construct as the measure attempts to balance days at home with other unintended consequences.

As far as meaningful differences, the Scientific Methods Panel members questions the presence of meaningful differences in performance and the use of measure for quality improvement purposes, and whether the measure could be used to identify differences in patient function or health-related quality of life. And as Dale has already mentioned, they did not reach a consensus on the validity criterion.

Co-Chair Bratzler: All right. Thank you, Kim. And we'll go to Katherine. Do you have any comments on validity, or Dr. Curry?

Member Curry: I think Kim has done an excellent summary, I don't have anything to add. Thanks, Kim.

Member Gray: Yes. This is Katherine. Nothing to add. It's got a lot of stuff in here that the Scientific Methods Panel pointed out. So lots for the whole group to discuss.

Co-Chair Bratzler: Thanks. So, Adam, I'll turn it over to you to kind of lead this discussion. There are a few questions in the chat, but see if other committee members have comments.

Co-Chair Thompson: Yeah, I don't see any questions in the chat room currently. We have some discussion points going on, but nothing that specifically relates to validity. We do have the question a little earlier about risk adjustment, but I'm not sure if that's what we're supposed to talk about.

Co-Chair Bratzler: I actually -- so I'm just going to comment. Kim, I know you've been doing major migration for many, many years also. I mean I do have some concerns about the risk adjustment. I find

social determinants very hard to capture, but incredibly meaningful.

And dual eligible is one important metric, but I don't think it's purely enough. Again, comes from bedside grounding, and others just knowing that the challenges, the complexity of placement of patients and other things. So I'm curious if others on the panel have questions about the risk adjustment and social determinants of a particular is what bothers me.

Member Glomb: Hey, Dale, this is Brendle. I was going to comment. Just when we're doing risk adjustments for our Medicare members, we're following the standard CMS protocols obviously, and there's a way where we're interacting with the provider to evaluate all existing diagnoses, making sure that those are up to date, severity, et cetera.

We also corporately look at the social determinates of health, and kind of, if you will, build upon the required risk adjustment. So through case management, we are picking up on all of those things. And they we're giving that individual patient an internal risk adjustment which adds that in.

Your point to, you know, can we consistently dig down on this, and certainly from the provide perspective, are you going to be able to do that. I suspect that most insurers are going to do what we do. And so I think good interactions between your discharge planners in the field and the case managers within the insurance entity will probably provide for that more accurate risk adjustment.

But it is -- I'd love to say it's a standard, but I don't know that a standard yet exists. I'd love to see -- I'd love to see us review a measure some time that looks at risk adjustment as part of SDOH.

Co-Chair Bratzler: Other comments?

Member Susman: One of the things that I've been concerned about is to -- let's take the most extreme

examples, the highest performers, the highest decile performers, the lowest decile performers, and drill down into those two groups. Are there meaningful differences in the populations they're serving, the location of their populations, the degree that social determinants of health may play a role? Or are there clearly differences in quality of care that are impacting days at home?

Without that, I have a real hard time, along with all the other questions and concerns that have been raised, to buy that this is a valid measure that can be enhanced by changes in quality of care.

Co-Chair Bratzler: Any other comments? Susannah, I know -- yeah. And so --

Member Glomb: Dale, can I cut back in?

Co-Chair Bratzler: Yes.

Member Glomb: Because we're in the validity section, and I had mentioned earlier I thought I had some general comments that probably spoke most to validity. And I think that it follows along with some of the things that are being voiced.

You know, I mean I'm going to add a little bit of sinister side to this as well. I think that there could be a lot of variability in these numbers that would represent this big, this very broad spread based on primary care physician motivation if you will.

CMS has obviously pushed nationally very hard and appropriately to build a value-based contracting and to make it a universal concept. We're not there yet, but we're getting closer and closer.

I worry that in those at most at-risk practices, i.e., capitated practices where their potential reimbursement or gain from a value-based contract is the greatest. That, you know, the higher the risk score, the greater the per member per month reimbursement. And that's in order to make the PCP

not incentivized to not spend money if you will.

When we start to look at the responsibility of the insurer, they're looking at the short-term care, skilled nursing facility, rehab facilities literally on an every other day, every third day basis in evaluating. So it's unlikely that there's going to be a lot of time spent in the non-nursing home long-term care setting that's going to be inappropriate or not medically necessary, so from the insurer's standpoint, even if the primary care provider is in a full at-risk or capitated sort of situation.

So I am concerned that the primary care physician, and I'm not talking about broadly, generally, or anything but a minority, would, you know, they would be incentivized to push the member to stay in a home setting as opposed to someplace where the expenditure is greater, thereby up-siding their contracts, if you get where I'm going with this. And I think that that adds a lot of variability when you get out of a tight little setting with a tight little definition of an ACO in this case.

And I do want to just mention that I did get some information back from our folks here in the State of Texas, and that really only single digit percentage of our Medicare patients are being cared for in an ACO group setting. So enough said.

Co-Chair Bratzler: Thanks. So I do know Susannah was going to be coming to you, and I wanted to learn a bit more about your testing. Obviously, I know you've been doing this for a long time, but how you guys are looking at social determinates, which was one of the big points of discussion.

Dr. Bernheim: Yeah. Thank you, Dale and everyone for such a good conversation, and I'll focus first on the performance piece. So while it's true that right now we did not have a sort of national standardized where I can build into a measure, you know, the difference between a patient who has food insecurity, or housing insecurity, or all of these other really

important components of someone's life, we were able to do some testing and -- that gets at some of these concepts I think in important ways.

So the first thing I'll say is that we spend a lot of time thinking about social determinants of health and guiding Medicare on appropriate variables to use. And their report on this actually comes out strongly in favor of dual eligibility being a surprisingly useful variable. It is imperfect. But for older adults, it represents lack of income and wealth in a way that I think correlates with other issues.

And so just a reminder that this measure does risk adjust for that as a way of capturing at least some of that signal and acknowledging that those can be challenging circumstances that will change an ACO's ability to keep patients at home.

We did also look at a number of other variables and looked at how well they correlated with the outcome. We looked at a measure of the composite, looking at a number of different factors in the neighborhood that a patient lives in. It's not a perfect marker for an individual patient, but we know that neighborhood is a powerful marker.

This is a SES index developed by AHRQ, it includes information around housing, employment, and education level. There's seven variables, I'm not going to name them all, but then we had poverty levels in the neighborhood. We found that that was less strongly correlated with the outcome of this measure. So we did not use that in the model. We also looked at some things about supply that came up earlier in rural and urban factors, and one other that I'm forgetting about.

So we tried to look broadly with the information that is available in all the Medicare patients, and found that dual eligibility was the best variable. One thing that often happens, and I'm sorry, my dog has joined me and is trying to join the conversation, so if you're hearing a little bit of scratching and panting in the

background, that is me, but I can't help it, so try to ignore her input.

The one thing I will say is people are often surprised when they see that the social risk factors that we can test don't have as strong a relationship with the outcome. And what we think is the reason for that is that this is a pretty thoroughly risk adjusted model.

So if you looked at some of these factors alone to see how much they individually might be associated with poorer outcomes, you might see a much stronger relationship. And patients who live in resource poor communities and have less social support often have greater numbers of co-morbidities, and so some of the adjustment that people expect to see is often actually accounted for. We've shown in the other measures basically we sort of add all of the other factors, these soak up a lot of the impact of the social risk factors.

So just to say we tested as broadly as we could. We often do not include social risk factors in our measures because there are competing concerns. But in this one, we, like you, thought it was an important variable to include. And we found the dual eligibility to be the best predictor of things that are available, and that would be consistent with what others have done.

I'm going to say just one word on the sinister thing because I appreciate that. I mean I think part of our job is all to -- this is to the comment that this measure could disincentivize good care. You know, I think it's important that we always think about that, and measures always can have unintended consequences.

You know, again we tried in developing this, and I will say that the SMP was initially confused by the approach we took, and so it contributed to a complicated conversation that I hope my introduction helped all of you to better understand it which is that this is an overall measure of days at home.

But that we wanted to ensure, at least in part, that care that led to less PACE circumstances or a higher tendency to transition people into long-term nursing care and at an excessive rate would be adjusted for, the excess days associated with those issues would be adjusted for this measures so that we are guarding somewhat against some of the more unintended consequences of a measure.

But, again, I think the evidence is pretty strong that there are things we can do and communities that do them well that keep patients at home. Some of our TEP members sit in a setting about PACE, and they have really shown how high the coordinated care can take very sick, older adults and allow them to stay at home much more.

Co-Chair Bratzler: All right. Thank you. Before we vote on validity, I'm going to see if there are any other comments by committee members. I've been reading through the chat. Anything else that jumps out at you, Adam, that we need to discuss?

Member Gray: This is Katherine. I was just wondering if we got the updated information from the CORE that applies to the validity discussion.

Co-Chair Thompson: Yeah. I believe that information was part of an Excel spreadsheet, right, and had the comments from the committee. There was a couple documents and thoughts from the measure developer.

Member Gray: But where are --

Co-Chair Thompson: It is possible for us to send that around again to folks?

Dr. Bernheim: What we saw was that it was tucked at the very, very end as a public comment, just so you know. It's a little bit hidden away.

Member Gray: Okay. So it was at the end of the full document?

Co-Chair Thompson: Yeah. At the very end.

Member Gray: Yeah. Okay. I got it then.

Co-Chair Bratzler: All right. Other questions or comments? Otherwise, we'll move to a vote on validity. Hearing none. So why don't we open the vote?

Ms. Kyle-Lion: Okay. Just give me one second. All righty. Voting is now open for Measure 3667 on validity. The options are A for high, B for moderate, C for low, or D for insufficient. We're at 17 votes, I believe we're waiting on just one more.

Co-Chair Bratzler: In the chat, one person did have to step away.

Ms. Kyle-Lion: Right. Still at 17.

Ms. Bal: Dale.

Ms. Kyle-Lion: Yeah.

Ms. Bal: Sorry, Gabby. Dale, we would still be looking for 18 because we had 19 last time with Jette joining. So we should still be looking for 18. But I think -- yeah, Gabby will give one more warning and then we'll just move forward with the 17.

Ms. Kyle-Lion: Still at that 17. We are now at 18. Okay. Voting is now closed for Measure 3667 on validity. There were zero votes for high. Three votes for moderate. Seven votes for low. And eight votes for insufficient. Therefore, the measure does not pass on validity.

Co-Chair Bratzler: All right. So I think that ends our conversation about this measure at this point if I incorrect, Paula and Poonam.

Ms. Farrell: That is correct.

Ms. Bal: That is correct. Yes.

Co-Chair Bratzler: Okay. All right. Well, so, Paula and

Poonam, and I'll turn it over to you at this point as we think about next steps here.

Ms. Farrell: All right, thank you. So it's now a little after noon Eastern Time so we're going to go ahead and take our scheduled 30 minute break. If you all could please return at 12:30 p.m. Eastern Time and we'll begin our discussion on our next measure at that point.

Co-Chair Bratzler: Thank you.

(Whereupon, the above-entitled matter went off the record at 12:05 p.m. and resumed at 12:33 p.m.)

Measure 3661

Ms. Farrell: All right. Welcome back, everyone. I hope you all enjoyed your break. We have two additional measures that we're going to review this afternoon. Our next measure for review will be 3661, Mismatch Repair or Microsatellite Instability Biomarker Testing Status in Colorectal Carcinoma, Endometrial, Gastroesophageal, or Small Bowel Carcinoma. So I'm going to turn it back over to our co-chair Dr. Bratzler to lead this discussion. Dr. Bratzler.

Co-Chair Bratzler: Yeah. Thank you. So I'll admit this measure took a little homework on my part to understand what we were, you know, going to be talking about. But it's a measure developed by the College of American Pathologists, and I believe they have a representative here that will give us a brief overview of the measure.

Dr. Bosci BOSCI: Yes, that's me. Hi, everybody. Can you hear me?

Co-Chair Bratzler: Yes.

Dr. Bosci: Okay. Thank you very much. So I'm Greg Bosci, I am an anatomic and clinical pathologist, and today I'm speaking on behalf of the College of American Pathologists.

CAP has a lot of experience creating and maintaining a spectrum of measures in their qualified clinical data registry that pathologists use for merit-based incentive payment system reporting, and we're asking NQF to endorse a very solid measure.

As I'm sure you know from the measure worksheet, this measures the percentage of surgical pathology reports for colorectal, endometrial, gastroesophageal, or small bowel carcinomas that address MMR or MSI status, or both. Because MMR or MSI status is important information about a patient's disease that isn't consistently captured, this measure is designed to identify higher quality practice amongst pathologists.

Regarding the importance to measure and report, looking at this measure and the quality action underlying it from a variety of angles, we see that a performance gap exists demonstrating that it's worthwhile to measure. And, you know, the basis for this quality action is very well established in the literature, and, honestly, that's a prerequisite for getting pathologists to use it in the first place. So as a matter of fact, CAP has seen that this is important based on many pathologists choosing to use and to report their performance on this measure.

With regard to scientific acceptability, we're really very happen with our testing results that demonstrate that we've got a measure that's valid and reliable. The outcome of the validity testing was particularly satisfying because it includes not only the perspective of pathologists, but also gastroenterologists and genetic counselors showing that there's a broad consensus that this measure can identify higher quality care.

The reliability testing established that this is a reliable measure. It scored very highly in the statistical analysis. And traditional score card feasibility revealed no significant issues which was no surprise as this measure is in use and pathologists are

successfully reporting on it.

So across the board, we have great confidence in this measure, and we brought it to you and think you'll find it suitable for endorsement. I guess since I don't know if I'll get to comment at the end, I'd also like to take this chance to appreciate everyone taking the time to thoughtfully consider the measure, committee members, NQF staff, and I guess lead discussants who will be taking it from here. And I'll be available to answer your questions, so don't hesitate if you need any clarifications. Thank you.

Co-Chair Bratzler: Yeah. Thank you very -- could you just perhaps help me just a little bit understand what this requires at the level of the pathologist to do this additional testing because when I read --

Dr. Bosci: Sure.

Co-Chair Bratzler: -- through the report, I noticed that, you know, it seems to be done consistently in academic centers, but maybe not in other places. So --

Dr. Bosci: Sure.

Co-Chair Bratzler: -- is it readily available technology something that any pathologist should be able to do?

Dr. Bosci: Yes. And so absolutely mismatch repair immunohistochemistry is done widely in pathology practices. Microsatellite instability testing is done less widely, but is widely available as a send-out procedure.

But I'd just point out that this measure doesn't necessitate a particular pathologist performing that testing. The key is that they address the status of those biomarkers in the report which is something that, you know, isn't required or demanded. But which, you know, based on our literature review, based on what we've seen, and is widely understood to be an important prognostic and predictive feature

of one's disease for these neoplasms.

And so I'm going to tell you, pathologists are a persnickety bunch, and if we dare to put down a measure that required them to do testing and, you know, that wasn't clearly indicated, they'd be all over us.

But we've crafted this and, you know, it's in use in a way that still satisfies the goal of sharing that information to avoid ,honestly, delays in care and duplicative testing for patients with these tumors in a way that's, you know, acceptable to pathologists and achieves that quality goal.

Co-Chair Bratzler: Great. Thank you. That's very, very helpful. So looks like our lead discussant is Karen Fields, and our secondary discussants are Jennifer Malin and Shelley Nasso. So, Karen.

Member Fields: Hi. All right. Thank you. And thank you, Greg. Greg's on our Cancer Measures Committee and I would have requested that he be one of the primary reviewers, but instead he's one of the measure developers. So it's good to see you today, Greg.

Dr. Bosci: Yeah. Nice to see you, too.

Member Fields: So as the description of the measure is that the measure seeks to measure the percentage of surgical pathology reports for various diseases, primary colorectal cancer, endometrial cancer, gastroesophageal cancer, or small bowel carcinoma either through evaluating primary biopsies or resection tissue, and contain the impression or conclusion of the recommendation for testing for MMR.

And there's several subsets of the MMR, including MLH1, MSH2, MSH6, and PMS2, or microsatellite instability MSI testing. And MSI testing, as Dr. Bosci indicated, is based on DNA-based testing status, and that's a send-out versus immunohistochemistry. I'm

going to refer to everything as MMR and MSI for the future.

The level of analysis is at the clinician semi colon group/practice level. And we'll spend some time discussing that later when we get to some of the sections. And this is a new measure. It's a process measure, and I'll discuss the evidence at the moment.

MMR and MSI testing of a sample are frequently needed to guide treatment decisions, particularly for patients who are being considered for checkpoint inhibitor therapy. In the absence of this testing, patients could be potentially treated with chemotherapy agents that they won't benefit from. And the testing is also a critical prognostic marker to determine the presence of a syndrome called Lynch syndrome which is an autosomal dominate genetic disorder that's associated with an increased risk for various cancers. So it has therapeutic and prognostic implications.

And pathologists are uniquely well-positioned at the time of signing out of the pathology report to detail the description of MMR/MSI testing for that sample. And referring physicians, oncologists depends on both the pathologist's interpretation and/or any recommendation for further testing in order to provide high quality patient care. So it is appropriate that targeting this measure towards pathologists at the time that they're signing up the specimens is done.

Just as a brief background for you, checkpoint inhibitors are a newer drug, a new drug called pembrolizumab, which was FDA approved in 2017, is a drug that is an immunotherapy that treats certain genetic markers particularly MMR and MSI. And this is an important new drug, and it's also important to note that this was one of the first cancer drugs that wasn't specific to an organ, or a particular part of the

body. But instead, it targets specific genetic markers.

I also will add that this drug is expensive and it is not without toxicities. Genetics testing is needed to identify which patients for whom this therapy might be effective. And this measure qualifies the rate of documentation of the specific genetic alterations that might make a patient more or less likely to respond to this cancer drug.

MSI is also a prognostic marker as we discussed for Lynch syndrome. Lynch syndrome is an autosomal dominate genetic disorder that puts patients at increased risk for various cancers, particularly colorectal cancer, but also endometrial cancers or certain kinds of brain tumors, et cetera.

The developers performed a systematic review of the literature. Initially, they identified a total of 6,642 potential studies that met the eligibility criteria for screening, and of these, 427 met the inclusion criteria and were continued on to a full text review. And from these, 103 articles were included for the systematic review of the literature.

And, subsequently, guidelines were developed from CAP, but the AMP, the Molecular Society as well as ASCO and Fight Colon Cancer also have developed these guidelines.

And there's three specific guidelines. The first one is in colorectal cancer, that patients should be measured for -- be considered for immune checkpoint blockade. And the description of measurements specifically focused on either MMR by IHC or MSI by PCR is noted.

The second guideline is in gastroesophageal and small bowel cancer patients, and in this recommendation, a certain kind of esophageal cancer is excluded, squamous cell cancer of the esophagus. And then in endometrial cancer, again, this is recommended testing.

So the systematic review was evaluated by a panel of technical experts who felt that each one of the guidelines was based on a strong recommendation based on medical evidence. And they felt that all three recommendations were accurate and carried large amounts of -- carried high benefits with only small harms. And in their evaluation, there was only one potential harm that was identified which was that PCR testing is slightly more technically challenging and has a higher cost than other kinds of testing.

So the relationship to clinical outcomes, the logic model indicates that the testing for these genetic alterations is critical to guide personalized treatment and assess the risk of cancer progression and development.

So my assessment is that the literature is strong to support the use of these genetic markers to determine whether or not a patient's a candidate for a specific type of therapy or at high risk for developing another cancer which would also recommend treatment for cancer prevention and follow up.

I also will note, and the investigator -- or the developers didn't note that the NCCN, which is the National Cancer Coalition Network, also has a recommendation that these genetic markers be evaluated in all patients.

So I think that the reviewers presented a good systematic review of the evidence. That the evidence was of high quality, quantity, and consistency. And the evidence that they presented was graded. So based on the algorithms for NQF, this measure meets the moderate rating for evidence.

And to summarize the pre-evaluation comments from the committee, most of the committee felt that the evidence appeared to be consistent, and no significant questions or concerns. Although, some concerns that this is a new type of measure for this committee to be measuring. So I'll open to any

questions, but before we move on, I'd ask my co-reviewers, Jennifer and Shelley, to comment.

Member Malin: Thanks, Karen. I guess just a couple things to add for additional context. So this is a measure that's similar to a measure that's been in place for I think several decades for breast cancer to ensure that the pathology reports for breast cancer specimen include estrogen receptor, progesterone receptor, and HER2 status which are, you know, important for making therapy decisions.

MMR and MSI are kind of, you know, analogous for colon cancer and endometrial, these other adenocarcinomas. You know, several other countries like Australia, for example, established universal testing I think maybe more than five years ago as Karen was mentioning because it's very important to identify individuals who carry a risk of hereditary cancer syndrome, and then can lead to genetic testing for the family.

It's important for identifying in people who have Stage 2 cancer, people who might not benefit from adjuvant chemotherapy, and then saves them the toxicity of having chemotherapy. And then now as Karen was mentioning, in people with advanced disease, it can help identify which patients would benefit from checkpoint inhibitors, this new class of immunotherapy drugs.

So kind of at, you know, at multiple different phases in patients with these cancers, this is really important information. And so just having it available on the pathology report so it's consistently available to all providers is really critical in leading to better outcomes for patients.

And I guess just the last thing that maybe when we talk about on that, you know, kind of variation, we're talking just about the variation and the performance of the measure itself.

But getting back to hereditary cancer syndromes, at

UnitedHealthcare when we've compared testing for the breast and ovarian cancer syndromes compared with genetic testing for the colorectal cancer syndromes, there's a tenfold difference even though they're prevalence in the population is about the same. So this is, you know, in general I would say identifying patients and families who can have these cancer risks prevented is a large unmet need.

Co-Chair Bratzler: Thanks, Jennifer. Shelley?

Member Fuld Nasso: Well, I think Dr. Fields and Dr. Malin really summarized it. I just want to say from the patient perspective, again, how important it is to have this testing to make sure that the treatments are the right treatments for the patient, and that the, you know, prognostic indicator for Lynch syndrome is identified.

But also just, you know, from a patient perspective, this is sort of out of their control. This is happening behind the scenes, and they have no real way to impact making sure that this happens for them.

So I think, you know, I think it's great that the College of American Pathologists is putting this in place and using this as a way to try to increase, you know, the adherence to that kind of testing because it's something that you, as a patient, cannot really advocate on your own behalf for. So making sure that it's being done behind the scene so that you get the right testing is really, really important.

Co-Chair Bratzler: Thank you. What a great overview of the measure. So, Adam, I didn't see anything in the chat. Are there other committee members that have questions for the committee, the lead discussants about importance which is the first thing we'll be voting on, and the evidence? You guys did such a great review. So I think, Paula, we can move on to voting on evidence.

Ms. Kyle-Lion: Okay, just give me one moment. All right. Voting is now open for Measure 3661 on

evidence. The options are A, high. B, moderate. C, low. Or D for insufficient. We are at 18 votes. I believe that there might be a 19th, so I'll just give it one more second.

Okay. I am still seeing only 18 votes. Voting is now closed for Measure 3661 on evidence. There were four votes for high. Sorry, 13 votes for moderate. And one vote for low. And zero votes for insufficient. Therefore -- give me one second to confirm.

Co-Chair Bratzler: Yeah, so it passes on --

Ms. Kyle-Lion: Yeah. Yes, thank you.

Co-Chair Bratzler: -- on evidence. So, Karen, can you go on and talk about performance gap?

Member Fields: So the CAP has an entity called the Pathology Quality Registry, I think I'm referring to that correctly. And they began this measure several years ago. As you heard, it's being reported -- it was developed and then is now reported as a MIPS measure.

And they saw that in the period between -- for the calendar year 2020, 33 CAP centers evaluated two of the existing measures, which were colorectal and endometrial cancer, and the average score that they saw was 78.3 presenting with a standard deviation of 20.9 points. So the scores ranged from 40 to 100 percent suggesting a wide variation and use of the reporting.

The following year, or most recently in 2021 with results through October, showed that the practices improved and the average score was 86.5. That's incomplete data for the year. And so there was evidence directly in a group of people or pathologists participating in a quality registry, which theoretically would have a higher level of motivation to do the reporting that there was a gap and a wide variability.

Also the developers cited multiple studies and

instances of low reporting -- or low measurement and reporting of these findings that ranged from a low of in the 20s up to the 50s. But consistently for all of the diseases, it was under 50 percent. So there seems to be a wide range in performance, and I think this is true clinically. Frequently, the clinicians have to go back and ask that these tests be done so that medical decisions can be made.

And the question regarding disparities, the developer stated that race, and ethnicity, gender, insurance, and/or socioeconomic status, and disability data was not readily available in laboratory information systems, and therefore, it wasn't captured by the Pathologists Quality Registry.

However, they did report in their review of literature that a 2020 study found that non-white -- or white, non-Hispanic patients were more likely to get testing done than black, non-Hispanic patients. And based on the payer, lower levels of testing were seen in patients with Medicare/Medicaid or no insurance compared to private insurance which only reported 47 percent in private insurance patients. So there is evidence in the literature that disparities are involved, but that's not a target of this measure to look at disparity data.

So I do believe that there is a significant gap in performance, and that now that this measure's being used for MIP reporting, that should also be associated with increasing motivation for performing these tests.

So my preliminary rating of the performance gap is high. And when I looked at the comments from the committee, five of the six responders agree that there was a performance gap. The sixth felt that there was a gap, but wondered if this gap warranted a national performance measure. So I'll ask Dr. Malin and Shelley to comment again.

Member Malin: Yeah, great summary. I don't know that really I have much to add to that. Thanks, Karen.

Co-Chair Bratzler: I think, Shelley, you are on mute. You said something but --

Member Fuld Nasso: Oh, sorry, I don't have anything else to add.

Co-Chair Bratzler: Okay. Adam, I'll turn it over to you, I did see a question or so in the chat.

Co-Chair Thompson: No, I don't think we had any.

Co-Chair Bratzler: The only one I see is, is there a reason why pathologists wouldn't send this off, difference on opinion about utility versus cost of the testing.

Member Fields: I'll comment. I think that just it's not been a standard that's been a recognized standard, and usually the clinicians, initially when all of this was available, the clinicians were prompting some of the requests for these drugs. And so I think that the move to have pathologists make this a standard component of reporting for these diseases will hopefully improve that.

And, yes, sending off PCR tests can be burdensome and more expensive. But I don't know that that's necessarily the motive. And I defer to others to comment on that.

Member Malin: Yeah. I mean I think as Karen says, I think the guidelines and consensus have moved from, you know, requesting testing in specific circumstances to seeing that there's enough individuals who need testing, and that it's kind of -- when it depends on getting requested at the point of care, there's just a lot of opportunities to miss getting it for the right patients.

And then, in addition, I think the cost of testing has come down substantially. So the cost of adding this testing is, you know, on the order of $100 to $200, which, you know, in the grand scheme of the cost of caring for these patients is pretty negligible. And it's,

you know, especially important, again, for all the downstream impacts that it can have on family members in terms of getting people to hereditary cancer testing.

Member Fields: To Jennifer's point also, NCCN incorporated this for all of these diseases fairly recently in the last several years, again, making that the standard of care of patients should be evaluated for these clinically. So I think it's a combination of the evolving data.

Co-Chair Thompson: You're on mute, Dale.

Co-Chair Bratzler: I was going to see real quickly if any other committee members had any questions or comments, and otherwise, we'll go to the vote on gap. There was a --

Co-Chair Thompson: Just seeing a question from Ester. So I'll pass it to the discussants and the supporting discussants. Is it reimbursed?

Member Malin: Yes. It's universally reimbursed.

Co-Chair Bratzler: Any other questions? All right. Let's move to a vote on gap.

Ms. Kyle-Lion: Voting is now open for Measure 3661 on performance gap. The options are A for high, B for moderate, C for low, or D for insufficient. We're at 15, 16. Okay. We're at 18 votes. Nineteen votes. I will go ahead and pull up the results. There were 11 votes for high. Eight votes for moderate. Zero votes for low. Zero votes for insufficient. Therefore, the measure passes on performance gap.

Co-Chair Bratzler: All right, thank you. So this measure did not go to a review panel because it wasn't deemed complex, so we'll be talking about both reliability and validity. So, Karen, I'll turn it back to you.

Member Fields: Thank you. So the numerator statement is that the surgical pathology reports that

contain the impression of, or the conclusion of, or recommendations for MMR testing or MSI testing.

So the test either has been done with immunohistochemistry, or recommended further testing as noted in the report. And I think that's clear and not unambiguous. And then versus the denominator which is all surgical pathology specimens for all the diseases.

And the main issue that I'd like to bring up while we discuss the results of the reliability testing are that the developers rated and reported that the measure will be measured at the clinician group practice level and the clinician individual level, and that is causing a little bit of confusion about the results of the testing.

I think that the data sources are well specified, and the methods for conducting audits to determine if these are appropriately reported are correct. There's limited exclusion criteria and they're appropriate which would be if there was already a known diagnosis of MMR positivity or MSI positivity in patients with Lynch syndrome or in a patient that might never be a candidate for this type of therapy.

The other exclusions aren't technically exclusions, but the specimens that actually don't contain tumor, and they describe what the criteria for that would be, are excluded from the denominator and the numerator. And so I didn't have any problems with the description of how the measurements and the studies would be done.

And so the preliminary rating from the staff was that the results were moderate at the individual level, but insufficient at the clinician group practice level for reliability testing.

The reliability testing, they looked at 51 clinicians in academic and private practices, and there was a total of 1,282 cases of the various diseases. The 51 volunteer data extractors were instructed to pull the

cases per the standards. And the developers conducted a signal to noise analysis from Rand using a beta binomial method which with the beta distribution defined by alpha and beta to calculate within and between provider variances.

The results, the mean reliability score was 0.96. Anything above 0.7 generally indicates a reliable measure. It appears to me that the data was reported and analyzed mainly at the individual scores. The developers did comment on the fact that there wouldn't be a difference because they're looking at the 51 different pathologists regardless of whether it's -- and if they redid the analysis based on group, there wouldn't be a different calculation that would be done.

So the committee also saw some of these concerns and some felt that the methodology was reliable for reliability testing was appropriate, and some noted that there was only a single reliability level testing performed, and that that was problematic.

So I think that the reliability testing was high at the individual level, and I actually would hope that we could ask the developers to comment on this and explain why there's a difference because otherwise, the criteria for NQF would mean that another calculation at the group level was needed.

Before I turn it over to any further discussion, I'll ask Dr. Malin and Shelley to comment.

Member Malin MALIN: I don't have anything else to add. Nice summary.

Co-Chair Bratzler: Shelley, do you have anything? Adam, we'll turn it over to you and I think --

(Simultaneous speaking.)

Co-Chair Thompson: You're on mute again, Dale.

Co-Chair Bratzler: Well, I don't think --

Co-Chair Thompson: Play with your mute button.

Co-Chair Bratzler: It says I'm unmuted.

Co-Chair Thompson: Still there. Oh, wait, try it again.

Ms. Bal: Adam, we can hear Dale.

Adam, can you hear us?

Co-Chair Thompson: I can you all, yeah.

Ms. Bal: Okay.

Co-Chair Bratzler: Okay. Well, now you should be able to hear me. They can.

Member Susman: Selective muteness.

Co-Chair Bratzler: So, Poonam, let Adam know to go ahead with the -- there was a question for the developer there.

Co-Chair Thompson: I can hear you. So the only question that sort of came up there was Karen sort of requested the measure developer to discuss a little bit about the lack of the testing at the group level.

Dr. Bosci: Yeah, sure. Happy to do that, and, you know, that's also a question I had. And so I'm glad that you brought that to our attention. You know, so the nature of that testing and the nature of the data that we collected, we did not know starting out what type of reliability we would demonstrate with this measure.

And so when our statistical consultants completed the analysis at the individual level, the results were so eye popping that, you know, aggregating it and repeating the analysis at the group level could only make the reliability better.

And since the reliability was -- I mean I don't know what the experience of this committee is with reliability testing, but on the Cancer Committee, we rarely see it so good. And that was the rationale for

not going further at that point. And, again, it was based on our statistical consultants informing us that reliability would only improve by aggregating it at the group level. So it was felt that this demonstrated both satisfactorily.

Co-Chair Thompson: Thank you.

Co-Chair Bratzler: Any other --

Co-Chair Thompson: Those are the questions, Dale.

Co-Chair Bratzler: Okay. Thanks. Any other committee members have any questions? So there is a question from Esther.

Co-Chair Thompson: Chat room from Esther. Is the method standard across sites? So I'll let the discussants take the lead on this question.

Member Fields FIELDS: So I'll comment. The methods were well described in the application, and there was not any ambiguity in the presentation of how the measures would be conducted. And so I thought that it was consistent and likely to be standardized across all the sites because there wasn't ambiguity in how to calculate the measures, and then what the measure itself defined.

Co-Chair Bratzler: All right. Any other questions at all from any committee members? So let's move ahead with the vote on reliability.

Ms. Bal: Dale, sorry, before we jump into the vote, I just wanted to make it clear what our options are for voting. Since this measure is specified for both levels of individual and group level, the developer has provided their rationale for why only individual is provided.

So if the committee feels like that rationale is agreed, they want to vote on both individual and group together, we can do that. If, you know, standing committee feels that they would like to vote on individual and group separately, we can do that as

well. So we just want to just let you know that there's multiple options.

If you're feeling that you are more likely to do individual versus group, then we probably should vote separately so the individual doesn't go down along with group. But if it seems like what the developer has provided in their rationale, it sounds great, makes sense from a methodological standpoint, we can vote on both together. But I did want to just make sure the standing committee understood there were some options here.

Co-Chair Bratzler: So this is my personal opinion, I actually agree with Greg that testing it at the group level probably would even increase reliability greater than doing it at the individual clinician level. But if there are other committee members that feel otherwise, that we should separate those votes, let us know now.

Member Susman: Do you need a motion to consider them together?

Co-Chair Bratzler: I guess --

Ms. Bal: The default is together. It would be a motion to vote on them separately.

Member Susman: Okay. Then I have no motion.

Member Fields: I --

Co-Chair Bratzler: Go ahead.

Member Fields: -- agree with some of the comments because it's one measure and whether it's segregated into small units or summarized, you know, in a summary group, I can't see that there would be a difference in reliability based on the data that was presented. But I also will be the first to confess that I devoted these massive calculations, so. I think Jennifer was going to comment.

Member Malin: I think I was just going to endorse

that we consider it as is. It's --

Co-Chair Thompson: There was a question in the chat room about just the guideline.

Ms. Bal: And I think that's for NQF staff. The comment was any concerns that the only individuals provided for the two. With NQF, you know, we really leave that decision up to the experts. So if you feel that the developer's rationale for only doing individual makes sense because it's being grouped up to the group level, then that's where we can go with it.

But, you know, we put insufficient initially because that's the starting point, and then you as the experts can help us determine if that rationale makes sense. So it is fine to except that rationale if you see that's fair.

Co-Chair Bratzler: So in the interest of moving us along, I want to see does anybody feel strongly that we should separate the vote on reliability into group level and individual, or just keep them together and do a single vote?

Member Gray: This is Katherine. I think, you know, it may be difficult to figure this out, but if, in fact, you have to have data in order for us to kind of know that this is true, we would be safe as to do it separately because we can see it for the individuals.

But you don't know within groups what that means that some people are not as likely to do it, or, you know, can't do it correctly for some reason, that, you know, it just means we don't really -- you know, we're kind of guessing. And so the safer thing would be to vote separately I think technically.

Co-Chair Bratzler: Any other comments? So Katherine has suggested maybe separately. I think we need a consensus though if we're going to do that. I guess my point, Katherine, is that because it's just a numerator/dominator process, the care measure,

the precision will get better at the larger group level for the metric versus the single individual clinician. But, you know, I wasn't a part of the expert panel.

Member Susman: No, I think this is just a basic statistical concept, and unless there are problems with aggregating to the group level, the reliability will undoubtedly go up.

Co-Chair Bratzler: All right. So Poonam has suggested we just vote separately just to be safe, which I'm fine with. Let's just move that along. Let's go ahead and get that done.

Ms. Kyle-Lion: Okay. So voting on Measure 3661 for reliability at the group practice level is now open. Your options are A for high, B for moderate, C for low, or D for insufficient. We're at 16 votes. I believe we're looking to be at 19. We are still at 16. Okay. I do see that there were two people that said they couldn't vote. Just do one last call.

Co-Chair Bratzler: I have had on occasion have to hit the reload to get it to open up on my phone, A little circular arrow, then once I do that, then it comes up.

Ms. Kyle-Lion: Okay, I'm seeing 17, I'll just give one more minute, one more second just to see if we can get the other two. Okay, we are still at 17, but for the sake of time, I think we can move forward. But I'll look to Poonam to confirm that.

Okay. One second. Okay, so the voting for Measure 3661 on reliability at the group practice level is now closed. There are six votes for high. Eleven votes for moderate. Zero votes for low. And zero votes for insufficient. Therefore, the measure passes at the group practice level on reliability. Do I move straight to the next level vote? Okay.

Co-Chair Bratzler: Yeah.

Ms. Kyle-Lion: Okay. Voting on Measure 3661 for reliability at the individual level is now open. The

options are A for high, B for moderate, C for low, or D for insufficient. Okay, we are at 17 votes. I'll just give one more second just in case we get the extra votes.

Ms. Bal: If anyone is having difficulty voting, feel free to chat Gabby your vote and -- oh, there are 19. Perfect.

Ms. Kyle-Lion: I see 19.

Ms. Bal: Just -- it came up as I was going to say it. Thank you.

Ms. Kyle-Lion: All righty, voting on Measure 3661 for reliability at the individual level is now closed. There were eight votes for high. Eleven votes for moderate. No votes for low. And no votes for insufficient. So, therefore, the measure passes on reliability at the individual level.

Co-Chair Bratzler: All right. Karen, I'll turn it back to you to talk about validity.

Member Fields: So the validity testing occurred through face validity from a group of 40 subject matter experts including pathologists, gastroenterologists, and genetic counselors, and overall four strongly agreed, 15 agreed, one did not agree, and -- or disagreed, and one strongly disagreed. Two strongly disagreed with the validity of the testing.

And I think that there -- as far as threats to validity, the exclusions were appropriate, risk adjustment was not done in this measure. The exclusions were well defined and appear to be appropriate. The comments from the committee were all fairly consistent that there was no real threats to validity, although, they noted that disparities and risk adjustment weren't evaluated.

And so I thought that there were no significant -- there was no evidence reported by the developers

that missing data was present, and no other serious issues were noted regarding the data. So I thought that the validity was rated as moderate per the TEP, and I thought that was an appropriate rating. And any comments from my co-reviewers?

So, sounds like no response.

Member Susman: Was there any further information on those who strongly disagreed? Seems odd that, you know, there are a couple outliers there.

Member Fields: Yeah. They didn't comment on the committee's comments. So I don't have an answer. I do think that the importance of their TEP was a variety of different kinds of clinicians that would be interfacing with that data. And, so.

Member Susman: Thanks, Karen.

Member Malin: Though, surprisingly, it doesn't sound like oncologists?

Member Fields: Correct.

Co-Chair Bratzler: All right. Any other comments from committee members, otherwise, we'll go to our validity vote? I think, Gabby, we can go ahead.

Ms. Kyle-Lion: Okay. All righty. Voting on Measure 3661 for validity is now open. The options are A for moderate, B for low, or C for insufficient. I see 17. Okay, seeing 19 votes. Voting is now closed for Measure 3661 on validity. There were 19 votes for moderate, zero votes for low, and zero votes for insufficient. Therefore, the measure passes on validity.

Co-Chair Bratzler: All right, thank you. So, Karen, let's go ahead and talk about feasibility.

Member Fields: Feasibility, all the data elements were generated or collected and used by healthcare professionals during -- or personnel during the provision of care. The data elements are in defined

fields in a combination of electronic sources, and the developer noted no feasibility issues.

The measure has already been used in 2021, and there were no obvious difficulties using and implementing the measure. And it appeared to me that this would be a feasible measure to implement, and there was evidence that they had been able to do it.

The only comment I'll always make whenever there's an audit-based measure is that it's an audit-based measure and requires auditing which is a burden. And I'll turn it over to Jennifer and Shelley if they have comments.

Member Malin: I mean I guess I'll just say that I would -- I mean it seems very feasible despite the need to extract the data. And typically, I think one often sees in these types of measures is that once it becomes kind of an expectation, then the reports are often kind of created in a synaptic fashion that actually greatly simplifies the data collection.

Member Fuld Nasso: I have nothing to add.

Co-Chair Bratzler: All right, thank you. I didn't see any questions come up in the chat. Are there any other questions or concerns about feasibility from the committee? All right, Gabby, let's open the vote on feasibility.

Ms. Kyle-Lion: Okay. All right. Voting is now open for Measure 3661 on feasibility. The options are A for high, B for moderate, C for low, or D for insufficient. Seeing 18 votes. Nineteen. Trying to pull up the results.

Voting is now closed on Measure 3661 for feasibility. There were six votes for high. Thirteen votes for moderate. Zero votes for low. And zero votes for insufficient. Therefore, the measure passes feasibility.

Co-Chair Bratzler: All right, thank you. Karen, talk about usability and use.

Member Fields: So with regard to use, we know that this measure's already been used for MIPS reporting since 2021. The measure is not being publically reported. With regard to is this meaningful to consumers or patients? Ultimately, a patient would want to know if he or she was eligible for a specific medication, and that that treatment was appropriately directed towards their care.

The patient would also want to know if they had a high-risk syndrome for which they were at risk for developing colorectal cancer or other malignancies because screening and prevention is important in these patients.

So I think that as far as use, this measure is a reasonable measure. It's in use. There's no measure regarding the results of the MIPS reporting. Although, the developers have previously reported that there were no problems implementing it. So this is a pass or no pass measure, and I would recommend passing. And I defer to Shelley and Jennifer.

Co-Chair Bratzler: Jennifer, you must be on mute. We're not hearing you.

Member Malin: Sorry, yes. I would agree that it's usable, that it's been in use, that broader use of it will help drive meaningful improvements in the quality of care.

Member Fuld Nasso: I agree with Dr. Fields' recommendation that it should be passed.

Co-Chair Bratzler: All right. Any other comments from anybody else? All right. Gabby, we'll move to the vote on use.

Ms. Kyle-Lion: Okay. Voting is now open for Measure 3661, sorry, on use. The options are A, pass, or B,

no pass. I'm seeing 18 votes. I believe we're just waiting on one more. I am seeing 19 votes. Just give me one moment to pull up the results. Voting is now closed for Measure 3661 on use. We have 19 votes for pass, and zero votes for no pass. Therefore, the measure passes on use.

Co-Chair Bratzler: All right, Karen, talk about usability.

Member Fields: Usability. So of note, the developers evaluated two time periods, one in 2020 and then more recently in 2021 and showed that the average score went from 78.3 to 86.5 suggesting that this measure would be usable to do performance improvement. And so, obviously, for all the reasons we've talked about, there is value to reporting these abnormalities.

There's no evidence that there would be any unintended consequences for patients. Although, we do need to remember that anytime we're doing genetics and genomics testing in patients, that there always is a small risk that could result in discrimination in employment or insurance. Although there's laws to protect, they aren't complete coverage.

I don't think that any of those reasons would be reasonable to not say that this is a usable measure for patients because personalized care, therapy directed appropriately, and risk factors for patients that are critical knowledge for improving their quality of care. So I would recommend that the measure's highly usable. And I defer to my colleagues for comments.

Member Malin: I would agree. I would say, too, I think the risk in terms of discrimination, this is a somatic mutation at this level, and you have to do further testing to determine whether it happened de novo, was actually inherited. So I think that risk is, you know, is not really kind of a true risk at this stage. It would be more the next stage.

And, you know, I think there's a tremendous amount of potential improvement in, you know, at a population level by increasing the identification of people with these hereditary cancer syndromes. You know, implementing colorectal cancer screening with colonoscopy at the recommended interval decreases the risk of death by about 80 percent in individuals with these syndromes. And it's vastly under identified at this point.

Member Fuld Nasso: Agree, nothing to add. Thank you.

Co-Chair Bratzler: Adam, I don't see anything in the chat. Any of the other committee members have questions for discussants? All right, Gabby, let's open voting on usability.

Ms. Kyle-Lion: Voting is now open for Measure 3661 on usability. The options are A for high, B for moderate, C for low, or D for insufficient. We're at 16 votes, looking for 19. We're at 18 votes, just waiting on one more. Still only seeing 18 votes. So I'll just give it another second.

All right. Pull up the results. Voting is now closed for Measure 3661 on usability. There were 11 votes for high, seven votes for moderate. Zero votes for low. And zero votes for insufficient. Therefore, the measure passes on usability.

Co-Chair Bratzler: All right. Karen, thank you for leading a great discussion. I think our last vote is on overall suitability for endorsement. Anything else to say, Karen?

Member Fields: No. Thank you for the opportunity, and thanks to my colleagues for contributing to the discussion.

Co-Chair Bratzler: Okay. I think, Gabby, we can go to the final vote.

Ms. Kyle-Lion: Okay. I'll just pull it up, give me one

moment. Okay. Voting is now open for Measure 3661 on overall suitability for endorsement. The options are A for yes or B for no. We're at 17 votes. I'm seeing 19. So I will pull up the results.

Voting is now closed for Measure 3661 on overall suitability for endorsement. There were 19 votes for yes and zero votes for no. Therefore, the standing committee recommends to endorse the measure. And I'll pass it back to you, Dale.

Co-Chair Bratzler: I think my hardest work is done, so I'm going to turn it over to Adam at this point.

Co-Chair Thompson: Thanks, Dale. Can you guys hear me okay? Can everyone hear me good? Awesome. I changed computers, so I just wanted to make sure. So thank you so much. And I just want to say kudos. I love our Primary Care Committee, don't get me wrong. But that was a beautiful presentation of that measure by our partners from the Cancer Committee. So thank you very much for doing that. It was really great to watch you all work.

Measure 3332

Co-Chair Thompson: Next up, we have our measure -- let me grab my piece of paper to get the number right -- 3332, Psychosocial Screening Using the Pediatric Symptom Checklist Tool PSC-Tool. We'd like to begin first, if we have our measure developers here, to give us a brief overview with three to five minutes.

Dr. Murphy: Hey, can you hear me okay?

Co-Chair Thompson: Sure can.

Dr. Murphy: So this is Michael Murphy, the co-developer of the Pediatric Symptom Checklist, and we are joined on the call today by my esteemed colleague, Dr. Michael Jellinek, a co-developer of the PSC and the original author of it. So Mike usually lets me do the talking in case I say something wrong, but

I'm sure he'll hop in if I screw up too badly. So thanks for making the time.

So I want to begin with a brief description of measure. Some people may not know it, the Pediatric Symptom Checklist is a one-page patient or parent-reported outcome measure that is administered during pediatric well-child visits to make sure that psychosocial functioning is assessed as a part of the visit. The official title of the NQF version of the PSC is Psychosocial Screening Using the Pediatric Symptom Checklist Tool, PSC-Tool.

And a brief description is it's the percentage of children from three to 17 who are seen for a pediatric well-child visit who have a pediatric symptom checklist tool administered as a component of that visit. So it's a process measure.

And before going any further, I want to thank all the committees. I feel like I'm going to get a medical doctor degree at the end of today. The discussions were enlightening, very inspiring actually. Great job. So thank you for that. And we also want to thank the NQF staff. We've worked with Poonam Bal in the past, and her help this time around was also great.

We also want to thank the PCCI Committee for adopting us. We were supposed to be reviewed about this time by the Behavioral Health Committee, but because of COVID, they canceled those meetings. And so we asked to be able to do our review this year anyway.

We've worked with Dr. Pincus and Dr. Susman and other Behavioral Health Committee members for the last two NQF reviews we've done, and it was great to know that there'd be somebody on the review who was familiar with the PSC and its history in NQF.

The PSC's history with NQF is actually quite long. The PSC was first endorsed by NQF in 2011, which was I think before the invention of writing, and it was re-

endorsed in two of its three subsequent major reviews. I've screwed something up about eight years ago and we lost our endorsement for a while, but we got it back.

I sat in on all of the meetings. Dr. Jellinek is also here to administer life-saving medication if I get too anxious again. These meetings are very stressful for developers. And I remember one particularly long and difficult review session where it looked like we were going down in flames, and first one pediatrician, and then another pediatrician, and then a third pediatrician chimed in and said, well, yeah, it may lack a few technical things, but most of us use the PSC all the time. We like it. Our parents like it. And for us, it's basic office equipment. So one of our mains claims for re-endorsement here is just the measure is used so widely as a routine measure in pediatrics.

A little data behind that is we did a quick count of six sites that we're aware of, and in those six sites alone since the last review in 2017, more than 2 million PSCs have been given in those six sites alone. And we have to reason to believe that the number is much, much higher than that. We get requests almost every day for the PSC to be included in electronic health records or practice set ups. So the PSC's very widely used.

Pediatric clinicians, as I implied before, find the PSC easy to administer, score, and interpret especially when it is available in electronic medical records as it is these days. Several very large studies, like 10 million cases, have shown that sites that require routine psychosocial screening have significant increases in outpatient mental health referrals and treatment -- almost done compared to sites that do not.

A series of smaller RCTs, and I think there are about four of them now, have shown significant decreases in symptoms for children who are screened positive

and receive treatment. Perhaps just as important is that other studies have shown that routine screening leads to an increase in discussions between clinicians and parents about children's psychosocial problems and what to do about them.

Not all positively screened children need therapy. And many positively screened children benefit more from guidance concerning exercise, mindfulness, or addressing social determinates of health like food insecurity. Routine screening with a PSC increases the number of conversations like this resulting in more at-risk children getting more help. So we continue to think that an increase in referrals is the most appropriate outcome to track in evaluating the PSC.

Finally, in the last year, the American Academy of Pediatrics has reaffirmed its decades-long recommendation that every child, every well-child visit include a brief assessment of psychosocial function. Every time every kid goes for a well-child visit, they should have some kind of assessment of their psychosocial functioning usually saying that it should be a brief standardized instrument of which the PSC is often mentioned by AAP sources.

To end with the pandemic, most studies we've seen show that rates of mental health problems seen in pediatrics are at new highs, and the PSC is being called upon even more often to help pediatricians to identify which kids need most -- which children most need help. How do you prioritize the kids who are anxious, or depressed, or getting lost?

We believe that NQF's endorsement of the PSC over more than a decade has played a role in the national increase in psychosocial screening, and it's our sincere hope that today's review will lead to a re-endorsement. Thank you.

Co-Chair Thompson: Thank you so much for that overview giving us history, a little bit of levity, and I hope we don't cause you too much anxiety. And if we

do, mindfulness and deep breaths, right, will get --

(Simultaneous speaking.)

Dr. Murphy: Thank you. Appreciate it.

Co-Chair Thompson: Great. Thanks so much. So just as a reminder as we begin our discussion, this is a maintenance measure. It is a process measure using claims data looking at the health plan level. Our leading discussant is Jeffery Susman and our supporting discussants are Raquel and Brendle.

So we'll begin first looking at evidence, and I would invite our lead discussant, Jeff, to begin with evidence particularly noting any new evidence that was submitted with this measure because it is a maintenance measure.

Member Susman: Yeah. Thank you very much, Michael, and the measure is one that, as been indicated, it's in the long-use in a variety of settings. And what is the evidence? Well, and here's where one has to ask how strict of an evidence find compelling.

Screening, in and of itself, probably isn't a patient-oriented outcome that anyone really cares about. It's what happens after the screening. And does the screening lead to some behaviors that ultimately will result in an intervention that improves a patient-oriented outcome that is of importance to them and their families.

So I think in the mental health field, increasingly we've moved from the simple act of screening to looking at the actual, at least intermediate outcomes that result from that screening. Referral might be one, but ideally, we would have even better outcome data around what are the actual impacts of an intervention that has resulted from the screening tool itself.

We also have had, I think, over the years qualms about specifying a single instrument as the only

method of screening. There are certainly aspects such as depression care where things like the PHQ-9 have been widely disseminated and used. But I think the field, you know, as a general, has some ambivalence about saying, well, you must use the Pediatric Symptom Checklist as opposed to screen for psychosocial problems affecting children. So those are two issues to think about.

The evidence that has been presented, the most strong randomized trials or pseudo-quasi experimental trials are linked to very robust interventions. And those interventions include, for example, the implementation of chronic care model within pediatric practices or system, the use of nurses to do an intervention. So it's always in tandem with some other strong intervention which has a strong evidence basis.

So, you know, you could come on either side of this. You could say, well, the screening itself would be say like taking a blood pressure, but not measuring the outcomes of treatment with blood pressure, or ultimately, what we hope we're preventing, strokes and heart disease and all the other things that hypertension is associated with.

On the other hand, the developers have presented information that, you know, provides a causal pathway, but it's not all together the most vigorous, robust pathway that one could imagine.

Yes, in states where this has been implemented like Massachusetts Medicaid, there appears to be a modest association with increased use of behavioral health services. Yes, in individual assessments, families, providers speak positively of the use of the instrument. So I'm a bit ambivalent about saying the evidence is there that a screening instrument alone has the impacts that we hope to desire, you know, improved patient outcomes.

So the question before us, you know, initially here is the vote on evidence where in the past it has been

seen to be high quality evidence. And whether we shouldn't be at least encouraging a move towards either a two-step process or the move to linking a screening measure with an actual intervention, or preferably, ultimately, a patient-oriented outcome. So let me stop there and turn over to Raquel and to our other reviewer, Bill.

Member Jeffers: So I'll just reiterate, Jeff, that was a perfect summary. I think it's amazing that since 2011 when this measure was first endorsed, and 2022, you know, we have obviously gained some ground in terms of increased screening. But really, you know, it would be amazing if we were advancing to more questions around integrated care models and, you know, to what extent because, you know, no provider wants to screen if there isn't someone to catch a positive screen and work collaboratively with families to get the child the right care and intervention.

However, I will say on the other hand that we don't want to lose ground where screening is happening using an evidence-based tool like the PSC, we also want to maintain that ground that we've gained. So I mean I guess I just agree with everything Jeff said which would have saved time if I just said that.

Member Susman: Bill?

Member Glomb: Yeah, this is Brendle. Yeah, I absolutely agree with the -- and as a pediatrician with the premise that the screen needs to occur with a well-child visit. I would echo, though, both Raquel and Dr. Susman, that I don't think this many years into it, I think the measure would benefit by evolving and that it demands at this point I think a second step.

Also claim space. So really just a new data point in addition to was a screen done, that takes some sort of result into effect, at least a completed referral as evidenced by a mental health/behavioral health claim from the system. I mean that's just a first step. That doesn't go into treatment and it doesn't go into

patient outcomes. But I think that that would really solidify the point of doing the screening at the primary care visit in the first place.

I also put a question out there in the chat, and perhaps you can answer this for me, Dr. Murphy, is there a CPT II code which is associated with a result, i.e., either a normal screen or a screen which raises concerns so that there could be yet another data point collected to add to the mix, if you will, screen done, screen abnormal, referral accomplished? To me that takes us down the road and really gives us a lot more information.

Member Susman: And just to be clear, a positive screen doesn't necessarily mean that you necessarily even have to refer. You know, I think watchful waiting, further evaluation at the PCP level.

Member Glomb: Absolutely.

Member Susman: The, you know, follow up, and coming in and saying, okay, you know, seems to look like there might be some depression going on here based on the subscales and the answer. All those would be reasonable. So crafting that measure is, well, not as straight forward as perhaps we could say sitting in the stands here.

It is a very interesting process here, and my ultimate plea, no matter what we do here to the developers, is to take it to the next step. We've been at this since 11. It's a wonderful tool. There's a wealth of data about it. It's been translated into many languages, you know, used in a bunch of different settings. The reality though is, you know, we're still a little bit in the dark ages here, and that's disquieting to me.

Member Glomb: I think to your concern, Dr. Susman, about the follow up, how do you prove that an intervention or an assessment has occurred post-screen, even if it stays in the pediatrician's office or as people are raising in the chat, you know, family practice office, we can't ignore the family practice

percentiles, I think that that would show up then as a visit with a primary diagnosis to use your example of depression.

So now the pediatrician -- because many pediatricians feel very comfortable in handling less complex, less severe mental illness or behavioral health issues. So I don't know what you all think about that.

Member Susman: No, I think, you know, again, it would take some thought clearly, and there are a lot of different ways. So the importance to measure, you know, I think it depends a little bit where you stand. This is a very broad-based screening tool.

There are a fair amount of, if you will, false positives that will end up not being, you know, grounds for referral or further intervention other than doing further assessment. But that's what the tool's meant to do. It isn't really a fault of the tool itself.

What we need to ask ourselves is does this really make sense in today's context, and how -- and I'm sorry for raising your blood pressure, that, you know, what are our expectations today? I don't think the evidence basis has changed about the integrity of the PCS. It's more about what do we do with it, and how do we know that we're influencing outcomes that are important to patients and their families?

Member Jeffers: Right. So just to build on that, I think that the question of today is more to what extent has the system built capacity for pediatricians and family practice physicians to counsel, manage, refer children with behavioral health issues and get them the right care, or provide the care in the office.

So I think that today's question is to move the system towards a better integration and increased capacity to manage behavioral health conditions in primary care settings.

Co-Chair Thompson: One thing, I just want to real

quick -- hold on real quick, I want to bring us back together and just do two things. One is a process point which is I want to keep us focused evaluating the measure we have, not the one that we might want. As much as I love this discussion and as a person very interested in behavioral health, I agree. But we want to keep it here talking about specifically about the evidence related to this measure.

So I do know there are some questions that were coming in from the chatroom, so I want to kind of turn to Dale and see, Dale, if there's anything that we should bring up now for the committee, or are all the questions related to the measure developer?

Co-Chair Bratzler: Yeah. So, I'm just looking through the chat, I think a lot of them have already been addressed. And, again, many of them addressed this comment that you just made, Adam, which we've kind of gotten off from the evaluation of this measure and we're talking about how it should be improved for the future which isn't our role today.

There was a question about, you know, was there a CPT II code associated with a normal screen versus others. And one other question was is related to family medicine which, you know, does -- family medicine doctors do provide a lot of pediatric care and does this metric include family medicine providers.

Co-Chair Thompson: Great. Thanks so much. Jeff.

Member Susman: As far as I understand, the code is essentially that the PSC was done, was performed. So it doesn't say whether it's positive or not. And my understanding is that in the Massachusetts Medicaid, that it's provider agnostic. But certainly the developer, if they have different information, could enlighten us.

Co-Chair Thompson: Thanks so much. Any further comments from the measure developer?

Dr. Jellinek: So are we allowed to speak?

Co-Chair Thompson: Oh, yes, go ahead. I'm sorry.

Dr. Jellinek: Okay. This is Mike Jellinek. So let me first say I share all of your concerns. The reason we -- and I developed the PSC is to encourage the development of a support system for primary care pediatricians. That was the whole goal.

When we started -- or when I started first, and then Mike joined me about 35 years ago, there was nothing in primary care for any pediatrician to use. And I was actually criticized for thinking about using it because there was no support system for it, so why should we even screen because there was no support system.

I took the opposite view, I felt that in the United States, if we identified kids correctly, there will be a support system. I wish it would have happened faster. It's gradually happening through population health, and, for example, at MGH, Mass General, where we do some of our work, they do have a social work backup system that integrates with the pediatric practices.

It's very variable across the country. There's some places where this gets picked up easily by a mental health support to the primary care pediatrician. In other places, we hear how desperate the primary care pediatricians are as they have to manage more problems themselves.

I have to say, although I'm old, I don't date back yet to the dark ages. And we have tried to look at outcomes both in terms of referral rates, and in our studies, it seems that the referral rate doubles. If you think of maybe 10 percent of kids have problems in a practice previous to using the PSC, 1 or 2 percent are identified. Afterwards, it's closer to 4 percent.

We don't want everybody referred. It's a first stage screening instrument. We want the pediatricians to

take care of as many of those problems as they want to within the office setting. We did actually try to find out what happens between the pediatrician and the patient, and referral in Chelsea, and in that study, we were able to look at the notes that the pediatricians wrote in their charts as well as the referral rate.

And we found that the referral rate did go up as expected, and that 70 percent of the pediatricians did note that they discussed the PS results during their office visit. So we felt pretty good about that. It did encourage discussion and it did encourage referral.

In our other studies, use of the PSC in schools, the PSC resulted in kids being sent to kind of CBT groups in Chile and that they had a CBT intervention with the children and the family, and the children that went to the intervention and the parents that went to the intervention seemed to benefit. So we have made an effort to find out what happens after screening.

Mike Murphy made comment on the code issue. My understanding is that we can look at the Medicaid referral code as an administrative billing measure, and that's the way we figured out that more children were referred in all the Medicaid clinics in Massachusetts after the PSC was used than before. But I leave that to Mike to comment on.

Dr. Murphy: Yeah. And I've got some specific studies I can cite that actually can answer some of these questions. But I don't know if people want to respond to Mike first, and then I'll throw in some numbers.

Co-Chair Thompson: Any committee members want to follow up on the previous comments? Okay, go ahead. You can give us some numbers, and then we'll turn it back to the discussion.

Dr. Murphy: Well I'm sitting here praying for those three pediatricians to come back in and, you know, channel them. But maybe Dr. Jellinek has done that. But, you know, I want to begin with the one I know we can answer the best, and it's what Mike just

alluded to.

So in the state of Massachusetts in 2007, so this statewide program has screened more than 5 million kids in the last 15 years. It's ongoing. It's survived COVID. It survived, you know, administration changes. They invented one of the most amazing systems. It's really worth looking into and I can send some cites.

But they said to the pediatricians, you have to screen and you have to bill for the screen. And the reason they did that is some really forward thinking administrator said if they bill for the screen, then we have evidence that they screened. And so in Medicaid data sets of several million cases, Karen Hacker and I got to work on those studies, too, we were able to see there was a bill for the screen, and then there was a follow up for whether there was a second code for whether it was a positive or a negative screen. And then we could check the service use afterwards.

So her published papers going back eight years show definitively that new kids are found, kids that are found get referred and receive mental health services. And then she topped that study, she got all of the state of California's Medicaid billing for 2007, 2008, and Massachusetts.

So she had 10 million cases of well-child visits, and showed definitively more than anybody can ever doubt that the rate of mental health service use went up, you know, doubled as Mike said, in Massachusetts and stayed absolutely level in California.

And that the services that were used were counseling and evaluation, not meds or in patient. So that's just one, you know, one of the things you said you'd like to see. We have that. And there's just no doubt that when you screen kids, more kids get seen.

And just one more point on that. So we've been working with these state Medicaid data sets for a

decade, and they're wonderful. But we didn't have such good data on middle class samples. And just in the past couple years, we worked with a network of 15 suburban pediatric practices that have phenomenally good data, and we found the same things. So we have lots of data that screening is feasible, and that it does lead to increased mental health services.

Oh, yeah, and one more question, one more point. Doing a lot of work lately with real-world pediatricians and finding that they refer to a lot of things. They referred to mindfulness. They referred to activity groups. They referred to sports. You know, pediatricians have been doing this forever, and this particular pediatric network, the pediatricians make a point of using lifestyle modification intervention.

So it's not just that all kids are referred to counseling. It's also that kids are referred to -- or maybe food security programs that pediatricians are using the positive scores as a reminder to look at the psychosocial functioning of the kids and to craft interventions that meet the needs of those kids.

Co-Chair Thompson: Great. Thanks so much.

Member Susman: I appreciate --

Co-Chair Thompson: Go ahead, Jeff.

Member Susman: -- what Michael and his colleague have said. It's right on. But still the measure is screening. It isn't screening and follow up. It isn't screening and referral. It isn't screening and documentation. It's screening, administering a screen. And it is what it is. I mean I think the committee has to decide if that is strong enough evidence given the varied uncertainties that happened thereafter.

Dr. Murphy: Can I raise one more point?

Member Susman: Yeah.

Co-Chair Thompson: Hold on one second. Let me kick it back to the committee here real quick, and then we'll come to the measure developer with questions again at the end one more time. So any other questions or comments from the committee related to evidence?

Member McCollister: I guess I would just add that -- I mean we've evaluated and approved a lot of process measures as opposed to outcomes measures. So like I don't understand why we would need to see a referral or some sort of an outcome. I mean the fact that pediatricians, you know, aren't doing this as consistently as they should be, particularly given the, you know, proliferation of mental health issues and psychosocial issues amongst children these days, I think that's a concern.

So if this brings that to their attention and incentivizes them to at least consider those things, then that to me would be a win. And, you know, as the evidence developed, perhaps future measures can be based on what happens after that. But right now, just incentivizing physicians and pediatricians to be able to measure this consistently over time as a way of catching issues, I think is important.

Co-Chair Thompson: Thanks so much, Anna. Any other questions or comments from the committee?

Member Pincus: This is Harold. Just I guess my question is, you know, I sort of going along with what Jeff was saying in that would we approve a measure at this point saying that that you actually took somebody's blood pressure, and that was it. It seems to me that --

Member Malin: Well, as a --

Co-Chair Bratzler: So I think that gets into the gap discussion.

Member Malin: Yeah.

Co-Chair Bratzler: Is everybody --

(Simultaneous speaking.)

Member Malin: That's what I was going to say as well.

Member Susman: I don't know that that's solely a gap issue at all. But --

Member Malin: And I think those screening measures are tapped out, right? Like if a measure's tapped out and you don't have any room to improve, then maybe it's not useful anymore. But I think that's a different issue on whether or not there's evidence to support screening in and of itself.

Co-Chair Bratzler: Yeah, I agree.

Member McCollister: I mean this is a pediatric psychosocial version of the measure that we all just approved unanimously. I mean that was as specific using, you know, genetic testing and technology to make sure that pathologists were testing for specific types of tumors and making sure that patients were, you know, considered for the appropriate therapy.

I mean this is a questionnaire version of that for something that's more complicated and difficult to measure than genetics at this point. So I mean -- and it's incredibly serious and we're in the midst of an epidemic of psychosocial issues amongst adolescents and teens. So, again, the fact that it's just a process measure may not be ideal, but there's very few things as it relates to the mental health of children these days that are ideal.

Co-Chair Thompson: Thanks so much, Anna.

Member Fuld Nasso: I would just add. I just think given that the toll the pandemic has taken on kids and their mental health, this is not the time to be backing off on something like this just because it doesn't go far enough. I think, yeah, we can all agree. And we've struggled with these same questions in the Cancer Standing Committee, you

can't have a plan of care for pain until you assess the pain first.

So you got to measure, you got to start with the screening, and then you need to do more. But like I just don't see now as the time to back off of something like this with as big of a toll as kids are dealing with from the pandemic.

I mean I'm just -- as a mom of three, I've seen it in my own kids and I just, you know, have been reading a lot about the toll it was taking on kids. And I think we still need this even though we can desire more for the future.

Co-Chair Thompson: Great. Thank you, Shelley. So I think what we're going to do is kind of wrap up this evidence discussion because this is a maintenance measure. We do have a couple options here that we want to take a look at.

One is we can decide to kind of not open it up for a vote, accept the previous evidence decision from the previous endorsement. If there is a member of the committee, however, who would like to bring it up for further discussion and a vote by this committee, one person can call that question and bring us to a vote. But if not, we can accept the current evidence and move to performance gap.

So if there is anyone who would like to speak up for a vote by this committee, and I see Jeff has asked for us to call the vote. So, Poonam, can we have additional discussion or do we just move directly to the vote?

Ms. Bal: It seems like we've had a pretty thorough discussion unless someone wants to bring up something else. I'll just emphasize to please vote on the measure as is, and not the measure we want, but on that merit. Thank you.

Co-Chair Thompson: Great. Thanks so much, Poonam. So I think we can go ahead and bring up the

vote then for evidence.

Ms. Kyle-Lion: Just give me one second. Okay, voting is now open for Measure 3332 on evidence. The options are A for moderate, B for low, or C for insufficient. We're currently at 16 votes. I believe we're looking to be at 19. I'm seeing 18 votes. I'll just hold a second longer to see if we get that 19. I'm seeing 19, so I'll go ahead and pull up the results.

Voting is now closed for Measure 3332 on evidence. There were 14 votes for moderate, five votes for low, and zero votes for insufficient. Therefore, I believe the measure passes. We'll turn it back to you, Adam.

Co-Chair Thompson: Great. Thank you so much. And now we'll move to performance gap. So, Jeff, if you'd to kick us off. I think Jeff is on mute, I'm just making sure other people can't hear.

Member Susman: Yeah.

Ms. Bal: Yeah, we're not hearing you, Jeff.

Member Susman: Thank you.

Ms. Bal: Oh, there you go.

Member Susman: Yeah. Thanks and sorry. So the performance of the PSC is clearly very variable. There's good data that supports that that was cited, and it ranges in one state 14.2 percent to 71.9 percent, a standard deviation of 12.4 percent.

There's a lot of other data that would suggest that even in those areas where screening has been part of the ongoing process, that there's still a relatively large standard deviation and certainly outliers on the low side. So I think there's clearly a performance gap as far as conducting the measure or not, or instrument or not.

Co-Chair Thompson: Thanks so much, Jeff. And now, Raquel.

Member Jeffers: Thanks. Definitely a performance gap. I just had a question. I was surprised to see that there were no significant gaps in screening by race, and I was just wondering if the developer at the appropriate time could address that question.

Co-Chair Thompson: Thanks so much, Raquel. We'll add that to the list for sure.

Member Jeffers: Thank you. Anything you would add, Brendle?

Member Glomb: I would just add that -- and there may be a Texas pediatrician in the group today, too. I'm not a general pediatrician so I don't do the screening myself. I do know that Texas Medicaid requires that the screening tool be done as part of what their -- the Texas brand of Bright Futures, it's called Texas Health Steps, and in order to be reimbursed for doing your well-child visit, you've got to have the screen done at the appropriate time.

So I think that's a really solid connection there, particularly for this measure. And I think perhaps that's why any racial or socioeconomic disparity may -- at least in this state might not exist because it is a requirement in the Medicaid program.

Co-Chair Thompson: Thanks so much. Opening it up for any questions or comments from the committee for the committee or the discussants related to gap.

Co-Chair Bratzler: Well, the only question I heard was about the racial disparities with respect to gap.

Member Susman: If we want to talk about, I mean there were no racial disparities found, and at least my thinking of this, I'll certainly leave it to the developers to comment, is that when you're mandating a screen like in Medicaid for payment, then you administer this instrument to everybody that comes by.

The gap in services probably exists when you start

looking at completed referrals and actual patient outcomes I would supposed. But we don't have that data, and nor should we require that.

Co-Chair Bratzler: Yeah, I would guess, this is something we published on in the past and the between provider gap may be very different than the within provider. So rolling out a measure in your practice, you just do it to everybody. So you don't identify racial gaps. It's the practices that aren't doing it and are they taking care of disadvantaged populations that you just can't see because you don't collect the data from them.

Member Jeffers: Right. As well as the kids who are not getting to their well-visits.

Member Susman: Right. Yeah.

Member Jeffers: For those who make it to the well-visit who are on Medicaid, they're getting screened.

Co-Chair Thompson: Great. Thank you. Did that address your question, Raquel, or did you have -- would like further clarification on it?

Member Jeffers: I think it's addressed unless the developers have anything to add.

Co-Chair Thompson: Okay. Developers, anything you would add to that?

Dr. Murphy: No. Obviously, a great question. We were delighted to find that there are no, you know, in the samples we've looked at, there haven't been any clear, large racial or ethnic disparities. But as Dale said and somebody else said too, this is so confounded with practice, it's a practice-level thing, so there's a much stronger practice affect than there is. But both in terms of the screening and actually in terms of referral, in a Medicaid system, you know, most of the kids are low SES and there are not any glaring disparities.

Co-Chair Thompson: Thanks so much. All right. I see

no other questions or comments coming in the chat room. Anything else before we move to our vote? All right. I think we can go ahead and bring up the vote for performance gap.

Ms. Kyle-Lion: Okay. Voting is now open for Measure 3332 on performance gap. The options are A for high, B for moderate, C for low, or D for insufficient. And, again, we're looking for 19 votes. I'm seeing 18 votes. I'll just give it another second. And now seeing 19 votes, so I will pull up the results.

Voting is now closed for Measure 3332 on performance gap. There were three votes for high. Sixteen votes for moderate. Zero votes for low, and zero votes for insufficient. Therefore, the measure passes on performance gap. I'll pass it back to Adam.

Co-Chair Thompson: Great. Thank you so much. Next up we have scientific acceptability beginning with reliability. Jeff?

Member Susman: Yeah, so there was appropriate evidence presented of reliability using the chart indication as a gold standard, reliability appears to be quite high and certainly well within the range of acceptable.

Co-Chair Thompson: Thanks so much. Anything to add, Raquel?

Member Jeffers: Nothing to add.

Co-Chair Thompson: Brendle?

Member Glomb: Nothing to add, thank you.

Co-Chair Thompson: Any comments from the committee? Any questions from the committee for our discussants? All right, I think we can move to our vote on reliability.

Ms. Kyle-Lion: Okay. Voting is now open for Measure 3332 on reliability. Your options are A for moderate, B for low, or C for insufficient. Again, we're looking

for 19 votes. Seeing 16 votes. We're at 18 votes so I'll just give it another second for that last one. Okay. I am seeing 19 votes, so I will pull up the results.

Voting is now closed for Measure 3332 on reliability. There were 18 votes for moderate, and one vote for low. And zero votes for insufficient. Therefore, the measure passes on reliability. Back to you.

Co-Chair Thompson: Great. Thank you so much. Next up, validity. Jeff.

Member Susman: Okay. So validity was tested at both the patient encounter level, and then at the accountable level. The patient encounter level, the validity appears to be very high. The look at the data from chart review and the actual coding had 91 percent inter-coder reliability. There seems to have been a kappa of 0.84. They have some new data available where there were two coders and there was 100 percent agreement.

So the data elements themselves, the fact that whether this was done or not, appears to be very high. And as we'll talk about more, increasingly, this is embedded within EHRs and it's pretty easy to find.

Co-Chair Thompson: Great. Thanks so much, Jeff. Raquel, anything to add?

Member Jeffers: No.

Co-Chair Thompson: No. Brendle?

Member Glomb: I just want to -- I'm sorry, I'm looking at our definition there, and I just -- the second part of validity testing is adequately identifying differences in quality. I guess if this is an all in or all fail, then administering the test, the screen is adequately identifying a difference in quality by the NQF definition. And the --

(Simultaneous speaking.)

Member Susman: Yeah. I mean the follow-on

analysis that was done looking at, in this sample at least, the link between performance of the screen and behavioral health service utilization was much less strong.

Member Glomb: Right.

Member Susman: And whether you consider it adequate or not, I think eye of the beholder. There had to be, if you will, some manipulation which I could go through if you'd like, but I think, suffice it to say, the more distal you get from the actual concept, the screening, to actual service, to improved outcomes, the less -- or more tenuous the link becomes which is, you know, what you'd sort of expect.

Co-Chair Thompson: Great. Thanks. Anything else, Brendle?

Member Glomb: No, sir. Thank you very much.

Co-Chair Thompson: No problem. All right, any other comments from the committee? Any questions from the committee for our discussants? All right, I think we can move to our vote.

Ms. Kyle-Lion: Okay. Voting is now open for Measure 3332 on validity. Your options are A for high, B for moderate, C for low, or D for insufficient. And we're looking for 19 votes here. I'm seeing 17 votes. Just waiting on one more now. Okay, we're at 19 votes. Just give me one moment to pull up the results.

Okay. Voting is now closed for Measure 3332 on validity. There were four votes for high. Twelve votes for moderate. Three votes for low. Zero votes for insufficient. Therefore, the measure passes on validity. Pass it back to you, Adam.

Co-Chair Thompson: Thanks so much. Next up, we have feasibility. And, Jeff.

Member Susman: Yeah. So, you know, the question is are these routinely generated data. Increasingly,

as Medicaid and other systems are requiring this, it's certainly feasible, it's demonstrated to be feasible. It does have to be integrated into the practice, or an EHR if you're going in that direction. It's not routinely captured, but certainly it can be. So I'd say the feasibility is moderate.

Co-Chair Thompson: Thanks, Jeff. Anything to add, Raquel? No. Brendle?

Member Glomb: I don't, thank you.

Co-Chair Thompson: Okay. Any comments from the committee? Any questions from the committee for the discussants? All right, I think we can move to the vote on feasibility.

Ms. Kyle-Lion: Okay. Voting is now open for Measure 3332 on feasibility. The options are A for high, B for moderate, C for low, or D for insufficient. And, again, we're looking for 19 votes. At the moment I as seeing 17. All right. I see 19 votes. Just give me one moment to pull up the results.

Voting is now closed for Measure 3332 on feasibility. There were four votes for high. Fifteen votes for moderate. Zero votes for low. And zero votes for insufficient. Therefore, this measure passes on feasibility. Pass it back to you, Adam.

Co-Chair Thompson: Thanks so much. Next up we have use and usability beginning with use, which is a must pass criteria for maintenance measures, just as a reminder for folks. So, Jeff.

Member Susman: Yeah. So, as we've already discussed, this is in use and has been widely supported from pediatrics. There is the efforts of the Commonwealth of Massachusetts. So I think by and by there's clearly use of this and uptake. And that we have less evidence that there's actually a whole bunch of performance improvement activities and what that might entail, but I think that's a relatively less concern, and there probably are data that just

aren't presented.

Co-Chair Thompson: Great. Thanks so much, Jeff. Anything to add Raquel?

Member Jeffers: No.

Co-Chair Thompson: Brendle?

Member Glomb: No. Potentially useful, yes.

Co-Chair Thompson: Great. Thanks so much. Any comments from the committee? Questions for the discussants? All right, I think we can go ahead and move to the vote for use.

Ms. Kyle-Lion: Voting is now open for Measure 3332 on use. The options are A for pass or B for no pass. Again, we're looking for 19 votes. Okay, we are at 19 votes, just give me -- to pull up the results. Voting is now closed for Measure 3332 on use. There are 19 votes for pass and zero votes for no pass. Therefore, the measure passes use. Pass it back to you, Adam.

Co-Chair Thompson: Thanks so much. Next up, usability. Jeff.

Member Susman: Yeah. So, you know, the extent which audiences could use this data and performance improvement accountability clearly this isn't being used. I think something that I would highlight as maybe a slight deficit is looking for unintended consequences or harms.

Clearly, uncovering psychosocial issues might have some stigma associated with them. There may be many cases where a positive screen engenders concerns among parents that are not justified because there really isn't so much going on. There could be a referral of lots and use of lots of resources that truly don't make a huge impact on children's wellbeing.

There could even be a focus away from the few things that are really important, say a substantial

depression, because of the multiplicity of domains that are being measured. And that said, yeah, I think it's okay.

Co-Chair Thompson: Thanks so much. Raquel, anything you would add?

Member Jeffers: I mean I think just to summarize what we've said before that it's a necessary, but ultimately non-sufficient measure. But necessary.

Co-Chair Thompson: Thanks so much, Raquel. Brendle?

Member Glomb: Yeah. Along with the unintended consequences, you know, here I could speak to some very, very rural areas where there are few resources to whom to refer these kids for a positive screen. And these resources are and can be easily overwhelmed if there is not more introspective analysis within the primary care office.

Co-Chair Thompson: Thanks so much, Brendle. Any further comments from the committee?

Member Jeffers: Do we need to discuss harmonization or no?

Co-Chair Thompson: That will come after this part of the vote when we'll talk about related and competing measures.

Member Jeffers: Got you.

Co-Chair Thompson: Great question though, thanks.

Member Susman: Will we have to vote on this?

Co-Chair Thompson: Yep. So we're good? Everybody ready? All right. Almost end of the day. Let's go ahead and pull up our vote on usability.

Ms. Kyle-Lion: Voting is now open for Measure 3332 on usability. The options are A for high, B for moderate, C for low, or D for insufficient. And, again,

we are looking for 19 votes on this. Just waiting on one more vote. Okay, we are at 19. One moment to pull up the results.

Voting is now closed on Measure 3332 on usability. There were four votes for high. Thirteen votes for moderate. Two votes for low. And zero votes for insufficient. Therefore, the measure passes on usability. Back to you, Adam.

Co-Chair Thompson: Great. Thanks so much. So the next step will actually be voting on overall suitability for endorsement, and then we'll move into the discussion of related and competing measures. So any further comments folks have before we move to that final vote? All right. I think we can go ahead and bring up that vote then.

Ms. Kyle-Lion: Just one second. Okay. Voting is now open for Measure 3332 on overall suitability for endorsement. The options are A for yes or B for no. And, again, we're looking for 19 votes. We are currently at 16 votes, just looking for three more. Okay. I am seeing 19 votes. So just give me one second to pull up the results.

Okay. Voting is now closed for Measure 3332 on overall suitability for endorsement. There were 17 votes for yes and two votes for no. Therefore, I believe -- let me just confirm with my team -- that the standing committee recommends endorsement of this measure. Pass it back to you, Adam.

Co-Chair Thompson: Thanks so much. So I first want to say thank you to our measure developers for joining us here today. I hope that was not too much anxiety for you. We appreciate you being here to respond to our questions and comments as they came up. So we appreciate it.

And I would say if there's anything to sum up from the committee, I definitely heard great idea, but we'd like to see the steps that follow it, too. So if there's anything you can do with your colleagues and the

other folks to kind of push that ball down the court, I think the committee, at least here on the primary care side, and I think I'm hearing it from our other folks, too, would love to see that measurement advance down the field a little bit. So thank you again for all the work.

Dr. Murphy: And we thank you for sensitive listening and we will try to respond to your great suggestions. Thank you.

Related and Competing Measures

Co-Chair Thompson: No problem. Thanks you all. So now I will, believe, if I'm doing this the right way, kicking it back to Paula to talk about related and competing measures.

Ms. Farrell: Yes, you are correct. Gabby, can we go ahead and show the slides again please?

Ms. Kyle-Lion: Yep. Just give me one second. Sorry.

Ms. Farrell: No worries.

Ms. Kyle-Lion: Okay. Should be pulling up now.

Ms. Farrell: Okay. Well, I'll go ahead and get started while she's pulling up those slides. So next we're going to review any identified related and competing measures to address harmonization. None of the measures had any competing measures. You can go ahead to the next slide, please, Gabby.

So after reviewing the measures, the committee can discuss harmonization and make any recommendations on harmonization because -- and you can go to the next slide, please, Gabby -- because Measure 3667 did not pass on validity, we will not review the measures that were related to this. And so we'll go on to the next slide.

Measure 3661 had no related measures, and so we'll go on to the next slide. And Measure 3332 has nine related measures. They're all listed here on this slide.

And the developer did advise in their submission that out of the nine related measures identified, four do not overlap with their measure. And the remaining five NQF-endorsed pediatric mental health measures relate to the Pediatric Symptom Checklist tool because they all involve depression and rely on the PHQ-9.

The Pediatric Symptom Checklist tool does not compete with these five adolescent depression measures because the PSC does not have the same target population. It has a much broader focus.

Also the developer advised that all five of the currently endorsed measures that use the PHQ-9 apply only to adolescents that are already diagnosed with depression. Whereas, in contrast, the target population for the Pediatric Symptom Checklist tool is children as well as adolescents, and in it includes 100 percent of both children and adolescents not justified to 10 percent of adolescents who were depressed.

So does the committee have any opinions that we should discuss and does the committee think the measure specifications for the related measures harmonize to the extent possible?

Member Jeffers: I have a question.

Ms. Farrell: Sure.

Member Jeffers: Is it, in fact, true that the PHQ-9, that the 0712, the depression assessment, the initial assessment of depression, I understand that almost all the other measures someone is already indicated or diagnosed with a problem. But isn't the PHQ-9 used initially to identify depression pre-diagnosis? Maybe somebody who's more clinically experienced than I could speak to that.

Member Susman: Yeah. You know, I think, again, it gets back to our discussion last time. I mean when we started out looking at mental health measures,

we would use a PHQ-9 for screening. And we would identify people who had depressive symptoms. And, you know, there was a fairly strong link between depressive symptoms identified with PHQ-9 at a certain level with actual major depressive disorder which we know has some certain negative outcomes.

I think the field moved on and started to say, well, gee, just screening isn't enough. We need to have some form of either active treatment, or, nowadays, actually showing remission. You know, ideally, that's what you want. You want a person to get active treatment and remission, so.

You know, specifically here, I don't think they're really competing for the same sort of purpose which is a general screen, and with the Pediatric Symptom Checklist for a board array of psychosocial issues, one of which includes depressive symptoms and actual major depression. So I don't see it as an important order. I don't know, does that answer your question, Raquel?

Member Jeffers: I think mostly, Jeff. My question is would you do both? Like would --

Member Susman: I mean if you're in a system and you've decided to make a real concerted effort to better treat depression, you would first do a screen, and you might decide to screen everybody. And then there would be some efforts to, okay, well, what do you do if someone has a positive screen?

That might be referral, it might be treatment by the provider who's done the screening. It might be a follow up would be appropriate to see, well, is this transient or ongoing. You know, did you do a screen for suicide thereafter, blah, blah, blah, blah.

So just like with the Pediatric Symptom Checklist, I mean doing the screen is the first step. The second, and third, and fourth are the real, in my mind, the important where the rubber meets the road. Does that help?

Member Glomb: I agree. I think, Jeff, too, I think in the pediatric population that pediatric depression often presents very differently if you will --

Member Susman: No.

Member Glomb: -- than classic depression as symptoms in adults. So I think that this was a, you know, the screen and then a follow-on screen if you will, to try and drill down, right?

Member Susman: Yeah. Yeah. Absolutely.

Member Jeffers: That's helpful, thanks.

Member Pincus: Also, I may be wrong, but I think 0712 might also include depression screening with a PHQ-9 and follow up.

Member Susman: You might be right. I don't remember.

Dr. Jellinek: This is Mike Jellinek, am I allowed to make a comment? Yes? I don't want to --

Ms. Farrell: Discussion should really be amongst the committee, but if the committee has finished their discussion, we can allow you to comment on this.

Dr. Jellinek: Let me know when I'm allowed.

Ms. Farrell: Okay. Does anyone else in the committee want to make any further comments or should we allow the developer to respond? All right, Michael, I think you can go ahead.

Co-Chair Thompson: Wait. Real quick. Sorry, I was having a problem unmuting. I don't know, like this isn't necessarily a question. It's just a comment as we think about like harmonizing and all of these competing measures.

Just an anecdotal slice. As a person with HIV because of all of these depression screening things, I was screened for depression annually almost 10 to 12

times because of all of these different ways people look at it.

So just something to think about that as you standardize this measure, people will try to like catch it at all the points. And I appreciate the well-visit on the previous measure. But just this is a lot going on here, so I just wanted to kind of put that out there. Like some of the unintended consequences are when we slice it in these very nuanced ways, sometimes on the patient side, we get all the questions without the nuance.

So not a criticism of this measure. Just a comment in general about when we say something super important, we also have to think about how many times, right, do you get it in a year before it becomes now over-care and it becomes a problem.

Member Jeffers: Thank you for that comment, Adam. I agree. That was the reason for my question. Are we doing this -- are we essentially screening twice? But I hear you guys, that this is a next-level question. That if someone screens positive, a clinician might decide -- a physician might decide to go on further explore depression specifically.

Dr. Murphy: I can answer a factual. I'm looking at the specification now for 0712, and I think as Dr. Susman said a few minutes ago, it says, "The percentage of patients 18 and older with a diagnosis of major depression who have a completed PHQ-9 during each applicable four-month period."

It's sort of what Adam was saying. But the intent is, okay, we diagnose them, but now let's see if they're getting better. So you're supposed to have repeat screenings. So that's what it is.

Member McCollister: Yeah. And that's my sense. I mean my sense, and please correct me if I'm wrong, is that this is a screening measure. It's a way to catch kids who have issues. Like, you know, when I see my psychiatrist, you know, especially now that we're in

telehealth, they give you the PHQ-9 every time, and at first I was really annoyed by it, and then it was kind of interesting to see the impact.

Anyway, it's become, you know, as somebody's who's a data nerd who's monitored my glucose forever, it's interesting to see that trend, and I think it's helpful. But like that is for somebody -- I mean I had diagnosis of major depressive disorder since my diabetes diagnosis.

So it's kind of a maintenance thing, where this is trying to catch kids who may go in just for their annual pediatric visit to see -- I mean to me this feels like the -- what felt like a random, strange question a few years ago when people started asking me, the doctor, if I had been, you know, if I felt safe at home. It's like a way of screening it and giving somebody the opportunity to, you know, catch it. So I see them as being functionally different.

Dr. Jellinek: So I just wanted to mention that it's relevant. This week online, Mike and I and our team published a paper where we gave the PSC to parents and -- I'm sorry, the -- yeah, the PSC to parents and the PHQ-9 to teenagers in primary care pediatric practice. And we did that with 5,000 of them. You know, actually looking for and trying to help pediatricians deal with suicidal ideation. So that particular study is relevant to this question of whether you could potentially use both instruments depending on what your goal was. And you can use both instruments.

And in that paper, we tried to advise pediatricians how to use them together. One for the adolescent which has probably better or different validity in terms of depressive symptoms, and one for psychosocial dysfunction where the parents may be more accurate.

And to put those things together in terms of trying to define and track kids who are at high risk for suicidal ideation or other behavioral abnormalities. So there

is a place for it together, although I think the measures are different.

Member Susman: I think for our purposes today, I think we've considered the related measures, but they ultimately probably play a slightly different role, and play often a collaborative role if you will, line up with what you just said.

Dr. Jellinek: You said it better than I did.

Member Susman: No. You're doing the heavy lifting.

Member Jeffers: I agree.

NQF Member and Public Comment

Ms. Farrell: Okay. Great. If no one has anything else to add, we will move on. If we could go to -- yes, thank you, Gabby. So now we're going to jump into letting NQF members and the public have an opportunity to comment. So if you are either a NQF member or are a part of the public and you wish to comment on the discussion today, please either raise your hand or you can put a comment in the chat. And we'll just pause here for about minute to make sure everyone will have enough time for anybody who would like to comment to be able to do that.

All right, I'm not hearing any comments, we'll move forward with the next slide. And I am going to turn this over to Oroma Igwe, our manager, to discuss next steps.

Next Steps

Ms. Igwe: Thank you, Paula. So major thank you to the standing committee, developers, NQF staff, and general public. For the live record, before we go right into this, I would like to make one minor correction on a statement that I made during the early part of the presentation.

Earlier I erroneously referenced one of our other great committees, the Geriatrics and Palliative Care

Committee. I just want to reiterate on the call that today, the Primary Care and Chronic Illness Committee with representation from the Cancer and Behavioral Health Committees convened to review three primary care designated measures. So, although, equally great in value, this is not the Geriatrics and Palliative Care Standing Committee.

Moving forward. Next steps. A reminder that today, 3667 did not proceed with the recommendation for endorsement. It did not pass on validity. So measures which are not recommended will still move on to public and NQF member comment, but the committee will not revote on the measures during post-comment unless the committee decides to reconsider on the submitted comments that may come through, or there is a formal reconsideration request from the developer. Measures 3661 and 3332 were recommended for endorsement.

So all of these measures will move into the draft report and staff will prepare that draft report for the detail and the committee's discussion and recommendations. This report will be put out for 30 days, and we will be welcoming public and NQF member comments.

The staff will compile those comments and prepare them in order for the developer to review them. The committee members will also be able to view those as well.

Now we do have a scheduled post-comment call. It's already on your calendars. And if you find a need to convene for that call based on the comments you see, we will certainly convene. If we do not find a need for that call, we will appropriately cancel it.

Staff will also incorporate comments and responses to the comments into the draft report in preparation for the consensus standards approval committee. This is sort of our final convening body for the endorsement process. Next slide, please.

So here is a timeline of those next steps. The committee successfully completed the full evaluation and will no longer need to convene for this follow-up meeting that you see on the screen. So NQF staff will cancel the February 17th follow-up call so you're aware in advance.

The draft report comment period will be March 25th to April 25th. The committee post-comment web meeting is scheduled for May 25th. The CSAC review will be late July, and the exact date will be determined. And the appeals period will be held from July 21st to August 19th. Next slide, please.

So here's a brief look at what is ahead for the next cycle, spring 2022. Intent to submit deadline was January 5th of this year. We did receive three measures, two of which are maintenance, and one of which is new. And none of which are complex. Next slide.

So, as you know, the project team can be primarily reached via email at primarycare@qualityforum.org, or by phone at 202-783-1300. Of course, to stay up to date on project updates, you are all welcome to visit our project web page. And for the committee members, materials are always available to you on your committee SharePoint site.

Again, I want to just say thank you and go to the next slide. And pause here to pass it on to my colleague Paula to ask if there are any questions, and also to address any closing remarks.

Ms. Farrell: Thank you, Oroma. Obviously, we all are at the end of our meeting, so I just wanted to provide one additional opportunity for anyone who would like to speak up or have anything else to say. Please let us know now, you have an opportunity to do that.

Member Glomb: I don't want to keep us any longer. This is Brendle. There was a question -- a concern earlier about us kind of talking about how to make a measure better, and I've just been around with NQF

now for a long time, along with Dale and others, and Jeff, and, you know, we have done that.

We have taken new measures and said, you know, this isn't going to pass. But, you know, X, Y, Z might be a good addition. We've also seen maintenance measures where we've made recommendations. And the next time we saw it, low and behold they had made some positive changes.

So I'm a big efficiency expert -- not expert, big efficiency person. I don't want to see too much mission creep either. But I do think that we have something to offer because we've gone through this really, really sexy, NQF scientific analysis of these measures. And I think we can recommend things as part of our deliberations. Am I wrong on that, is my question?

Member Susman: Gee, I hope not.

Ms. Bal: Paula, did you want me to jump in?

Ms. Farrell: Sure, that's be great.

Ms. Bal: So, yes, you're definitely able to make recommendations on improvements. I think the only emphasis is that we should always vote on a measure as is, and make that decision not with the idea that we would like this better measure.

But in, you know, as you said, for the sake of efficiency and innovation, we should always make comments about in the future we'd like to see this, this type of measure, this improvement, this change. It's perfectly within the confines of the standing committee to make those recommendations as long as the future state does not impact your review of the current state.

Member McCollister: And, Poonam, how does that get registered? Like the -- because I know that I, you know, occasionally get a bit of a soapbox about one measure or another, and say let's do this. And

actually, I've seen some of the stuff come back up in future cycles which is, you know, gratifying in ultra-nerdy way.

But I mean like how -- is there any kind of formal process through which this is submitted to CMS? So if like, you know, the committee that voted this way for this measure, but what they think would be most helpful is a measure that, you know, for instance, like the first measure that we evaluated today, and issues with the specific design of it, but the issue that it was trying to get at was incredibly important.

So I can sort of riff on different ways that we might get to that particular issue, and I think those kinds of things should be, you know, requested for development by CMS.

Ms. Bal: Yeah. So we do update, including the report, any improvements that were suggested by the standing committee so we have that documented for future use. So when this measure comes back for maintenance review, or if it, you know, didn't pass and comes back for future reiteration, we do have that documented that, you know, there were certain suggestions of how to further improve this measure, or other gap areas. So that's part of the report.

In terms of necessarily making recommendations to CMS, you know, here at NQF, our role, especially for endorsement, is just to, you know, universally say these are what's good measures, and to more of the universe of developers and stewards, this is what we would like to see in the future. So no direct message to CMS, but there is a more broad message to the broader community of developers and stewards.

Member Susman: You know, I think one of the things NQF can be though is a voice to the field about enhancing evidence, the quality of evidence, the ties to patient-oriented outcomes that really matter. And not be simply satisfied with process measures that are not tightly linked to important outcomes.

I think that is a consistent concern or challenge for the field to continue to move forward. You know, the fact that I might get my A1C done is great. The fact that my A1C is in a good range is better. The fact that I don't go blind, don't have to go on dialysis, that's better yet, you know?

So I think that's where a voice, and I know it doesn't come necessarily at the committee level, but at the board, at CSAC, I think one of the things we need to do as an organization is to push the field forward. So off my soapbox.

Ms. Bal: We definitely agree with you, Jeff. We want to be moving things forward. I think there's definitely, you know, even within the standing committees, we make those recommendations to move towards those outcomes, closer to the outcomes if nothing else.

And even more broadly there, more conceptual projects that NQF works on where we also convey a similar message about this is what we want to see, this is what the future measurement is, and here's what, you know, we should be working towards that future as much as we can. So we, at NQF, definitely agree with you on moving towards those outcomes and making sure that we're shifting that way.

Ms. Farrell: Okay. Any other thoughts or questions?

Member Gray: This is Katherine. I just wanted to say for the developers for the PSC, that I think it was the most well-documented submission I've ever seen for a process document -- or process measure, sorry.

Dr. Murphy: Thank you.

Adjourn

Ms. Farrell: All right. Well, that is the end of our meeting. I'd like to thank our standing committee, our measure developers, NQF members, and the public for their participation. I will also thank you so

much for our co-chairs, Dr. Bratzler and Mr. Thomas, for their work and leading this meeting. And thank you for everyone to joining us, and enjoy your evening.

(Whereupon, the above-entitled matter went off the record at 3:07 p.m.)