NATIONAL QUALITY FORUM

Moderator: Kim Patterson
January 31, 2020
2:00 pm ET

Harold Pincus: One thing I would add is just, in terms of, if this measure is endorsed and comes - (is there a way) to sort of signal that, when it comes up for reexamination, that this issue be looked at, to see whether in fact there has been a change in terms of the, you know, the proportion of times that the risk of harm is actually acknowledged and identified, to see whether that sort of hypothesis that was just made actually is true.

(Nicole): (Harold) this is (Nicole) from NQF, from the staff side of things, it's something that we can capture as part of the committee's conversation and recommendation to take another look at upon maintenance review, so that we make sure that we bring the information for a discussion, if this measure does get into us and then goes through maintenance.

Harold Pincus: Okay, that's good. Other comments on validity? Okay. So I guess we're ready to vote?

(Hanna): Voting is now open for measure 3539E via voting on validity. Your options are A, moderate; B, low; and C, insufficient.

Woman 1: Just wanted to make a note for the voting categories for this one. You'll see it only goes up to moderate as the highest possible vote. That's due to the developer providing face validity as the type of validity testing. So, a passing vote in this case would be moderate.

(Hanna): Voting is now closed for validity on measure 3539E. We have 13 votes for moderate, 7 votes for low, and zero for insufficient. This measure passes on validity.

Harold Pincus: Okay. So now we move to feasibility.

Woman: Yes, correct.

(Mike): Yes, I'll take that one.

Harold Pincus: Okay. Mike.

(Mike): Yes. And then, so the comments from the group really revolved around this same issue we're talking about. One of the reviewers identified all data elements are in EHR or e-prescribing system, but a number of other responses really had to do with the item we just discussed, whether there's structured data in the EHR that could be pulled. And if it's not, it makes it difficult to pull. So those were the comments about feasibility.

Harold Pincus: Any other comments about feasibility from the committee?

Okay. So I guess we can vote on feasibility.

(Hanna): Voting for feasibility is now open for measure 3539E. Options are A for high; B for moderate; C for low; and D for insufficient.

Voting is now closed for feasibility on measure 3539E. We have 3 votes for high, 11 for moderate, 6 for low, and zero for insufficient. This measure passes on feasibility.

Woman: I think next we can move on to use.

Harold Pincus: (Lisa), (Michael)?

(Lisa): So this is (Lisa), and (unintelligible) together, if that's okay. Use, the data has potential (unintelligible) what you can conclude, but the comments by the committee members hadn't been used yet. The usability piece, there may be - there's a little bit more concern about the usability in terms of concerns of unintended consequences if prescribers decide not to prescribe the antipsychotics when in fact they really could be useful for the patient. Also that they may be using other medications, instead of antipsychotics (unintelligible) with other intended consequences. But there is, you know, obviously potential that this measure could be helpful in decreasing the use of the antipsychotics (unintelligible).

(Michael): Yes, thanks, (Lisa). I would agree. I had the same view on the usability, that it's really around the unintended consequences if somebody sticks to the measure, they may not actually treat the patient correctly, which I've seen happen, specifically with this kind of case, with an (unintelligible) person. So that was really an issue I think we just need to be aware of there could be real unintended consequences with this one.

Harold Pincus: Other comments from the committee?

So I presume there's no data to bear on whether those unintended consequences have occurred in certain circumstances where there's been efforts to reduce antipsychotic use.

Andrew Sperling: This is Andrew Sperling (unintelligible). There actually is, but not in the inpatient acute - acute inpatient setting. There's lots of - there's quite a bit of the literature about what has been done since CMS imposed the star ratings with the standards relative to inappropriate off-label prescribing of antipsychotics in nursing homes, you know, where we had (unintelligible) that would be used as (unintelligible) agents, and some of the unintended consequences there and some concerns around, for example, (unintelligible) homes with a history of (unintelligible) things like that, you know, haven't been able to get a nursing home bed out of concern that they'll be forced to use antipsychotics, in some cases patients who are prescribed antipsychotics in the community and that needed to go into a nursing home, that there were problems finding a bed because the nursing home was reluctant to take them because they, you know, they might - and again, the nursing home is - the building, the facility is penalized for the basis, you know, on the basis of that inappropriate prescribing financially with the star ratings, potentially financially.

So there are a few things out in the literature about the experience as these standards have now been in place for nursing homes for nearly a decade.

Harold Pincus: Uh-huh.

(Michael): But it's very, very different when someone, you know, someone's going in a nursing home potentially for a long-term stay versus someone who's in an acute inpatient unit, say, for, you know, chest pain or something like that

where they're not going to be there for a month, they're going to be there for three days. So, very different (unintelligible) nursing homes.

Harold Pincus: Anybody else have knowledge of other sort of data about the potential risk of these adverse consequences?

David Einzig: Yes, this is David Einzig. My perspective is I'm concerned about treating the label instead of treating the person. And it takes the judgment out of the clinician who's getting to know the patient and circumstance. It may alter it based off of a label instead of doing what really may be helpful for the patient.

Harold Pincus: Other comments from the committee?

Man: I guess I'll just (unintelligible) just like, you know, there's the last couple of comments were sort of about the potential of under-utilization of these meds, which I think is possible. We should keep in mind, there's almost certainly some over-utilization and we're going to need the balance of that.

Harold Pincus: Okay. So I guess we're ready to vote on use and usability?

Raquel Mazon Jeffers: This is Raquel. I just have a question. I know I'm joining this conversation late. But the measure is constructed that any use (unintelligible) any use of antipsychotic. Is there a way to measure inappropriate use of antipsychotics? Because isn't that what we're really trying to measure?

Man: Yes, I think the default is that it's inappropriate and so the exclusions, you know, sort of are dealing with the sort of appropriate uses. I mean that…

Man: Yes. And so remember…

Man: …framework for this.

Man: And so remember, in this context, remember in this context we're talking about (unintelligible) you're excluding people that (unintelligible) were on these meds prior to hospitalization, which was intended to pick up some of the appropriate use (unintelligible) and there are other exclusions like risk of harm to self or others. And so it's sort of a, Raquel, they've tried it, to take out, to identify several of those things that ought to be appropriate use of them and try to leave the (unintelligible) that might be an inappropriate use.

Raquel Mazon Jeffers: Right. Okay. Thank you for that clarification.

Harold Pincus: Other comments about use or usability?

Okay, I think we're ready to vote on use.

(Hanna): Voting is now open for use on measure 3539E. Options are A for pass; B for no pass.

Voting is now closed for use on measure 3539E. We have 18 votes for pass, 2 votes for no pass. This measure passes on use.

Harold Pincus: And on usability, any further comments on usability?

Okay. If not, we can vote on usability.

(Hanna): Voting is now open for usability on measure 3539E. Options are A for high; B for moderate; C for low; and D for insufficient.

Voting is now closed on usability for measure 3539E. We have 1 vote for high, 14 for moderate, 5 for low, and zero for insufficient. This measure passes on usability.

Woman: So we will now move forward to the overall suitability for endorsement vote. So, want to open it up to see if there's any comments before we do so.

Harold Pincus: Any comments for or against?

I guess we're ready to vote.

(Hanna): Voting is now open for overall suitability for endorsement on measure 3539E. Options are A for yes; B for no.

Voting is now closed for overall suitability for endorsement. We have 19 votes for yes, 1 vote for no. This measure passes.

Harold Pincus: So, now moving on to measure 3541.

Woman: That's correct. Harold, it's still you.

Harold Pincus: Okay. So, why don't we get an introduction to this from the lead reviewers, (Danita) and (Larry)?

Man: Okay.

Man: Harold, do you want the developer first?

Harold Pincus: Oh yes, first the developers. I'm sorry. Rushing ahead.

Ben Shirley: Hey. Can everybody hear me all right?

Harold Pincus: Yes. This is the Pharmacy Quality Alliance.

Ben Shirley: Yes, this is Ben with PQA.

Harold Pincus: Yes. So, why don't you go ahead?

Ben Shirley: Sure. Good afternoon everyone. My name is Ben Shirley, Associate Director of Performance Measurement at the Pharmacy Quality Alliance. And on behalf of our team, I'm pleased to present the annual monitoring for persons on long-term opioid therapy, or AMO measure, for NQF endorsement consideration.

Current evidence suggest that the majority of opioid fatalities originate from opioids prescribed within practice guidelines. This underscores the need for clinicians to have better information in the care and management of individuals on long-term opioid therapy to reduce adverse drug events. Furthermore, studies have shown that no set of characteristics sufficiently identifies patients at risk of opioid misuse or opioid use disorder. And positive drug screens are common in patients that display no external evidence of aberrant behavior.

One of the crucial evidence-based tools providers have to understand their patient's long-term opioid use is routine drug testing, which helps prescribing clinicians understand risk levels as part of the larger risk evaluation and minimization program. Annual monitoring of individuals on long-term opioid therapy in the form of a drug test is supported by clinical guidelines and is an accepted standard of care that empowers clinicians to identify patients that are not adherent to their therapy and at risk for adverse drug events or diversion.

Drug test results can influence referrals for substance use disorder, prompt education on the risk of potential drug-to-drug interactions, or highlight the need to change the treatment regimen when opioids are needed but not used. The AMO measure specified at the health plan level captures the percentage of individuals on long-term opioid therapy who failed to receive at least one drug test during the measurement year. Long-term opioid therapy is defined as individuals with at least 90 days' supply for any opioids in the measurement year, and individuals in hospice or with cancer are excluded.

The measure was developed in conjunction with a technical expert panel that provided significant input throughout the development process, and unanimously found the measure to have face validity. The measure is specified using administrative claims that are readily available and is highly feasible. Testing revealed both sufficient reliability as well as significant room for improvement and plan and product level variation, with nearly 70% of individuals, and the denominator are a sample of qualified health plans, failing to receive a drug test.

This measure is intended for use in the quality rating system for qualified health plans sold on the health insurance exchanges, but may be used in other programs in the future, with additional testing as appropriate.

The AMO measure represents a feasible, impactful, actionable and evidence-based measure that will drive improvement in a critical area's need for the health system. By encouraging routine testing, providers will be empowered with better information on their patients undergoing long-term opioid therapy, ultimately leading to safer patients and better outcomes.

So, thank you, and we look forward to the discussion today.

Harold Pincus: So, now, (Danita) and (Larry), you want to give your response and summary?

(Danita): Sure. (Larry), whichever way you want to go. Do you want to lead and I'll follow up, or whichever way you want.

(Larry): Okay. I mean that's fine and (unintelligible) thank you. So that was a very good summary. Basically talked a lot about what we're going to talk about.

So in this measure, which is a process measure based on claims data from a plan level, the numerator is individuals on long-term opioid therapy who have not received a drug test during their measurement (year). And the denominator is individuals 18 years of age and older who were prescribed long-term opioid therapy during the measurement year. As you heard, the exclusions are diagnosis of cancer of any kind during the measurement year, and hospice care at any time.

And so I think the first part (unintelligible) important to measure and report. There was a systematic review. Just looking at my notes here. A logic model was presented, depicted that drug testing would decrease the number of emergency department visits, hospitalizations and fatalities. There was a systematic review and grading of the body of empirical evidence. There were five evidence-based clinical practice guidelines, all of which included a myriad of studies which included randomized trials, observational studies, retrospective data analyses, and meta-analyses. They all articulated that drug testing would be beneficial (in approving) care. And they applied directly to desired outcome, at least in moderate fashion.

(Danita), anything you want to add to that?

(Danita): No. That was pretty complete.

(Larry): Thank you.

Harold Pincus: Do any of the other reviewers want to comment overall?

Raquel Mazon Jeffers: This is Raquel.

Harold Pincus: Yes.

Raquel Mazon Jeffers: Long-term opioid therapy is defined as how?

(Larry): Ninety days.

Raquel Mazon Jeffers: Uh-huh.

(Danita): So, Raquel, I think that is going to come up when we talk about reliability and validity. There's several people who have commented and I put forth also some questions on the definition. Do we discuss this now or wait for reliability and validity?

Peter Briss: Why don't we wait for reliability and validity?

I had a question, is, why did they choose, this is really for the developer, why did you choose one time per year? Was there some evidence that once is sort of a minimal threshold or was there some - is there some evidence that some kind of benefit with testing beyond that?

Ben Shirley: Hi. This is Ben. This is the developer. So our decision to use the one-time annual was based on some of those clinical guidelines. The CDC does

recommend annual testing. The AACC recommends one to two times per year, depending on risk levels. The annual approach was something we did discuss with the technical expert panel. And I think ultimately we did decide on the once per year as a reasonable threshold.

(Danita): And Peter, if I can add, when we talk about performance gap, you can see, even with just once per year, there's such a huge gap already that for CDC to just (unintelligible) minimum might be a good start.

Harold Pincus: So, do we want to talk about sort of - it sounds like we've already discussed evidence in the initial introduction, are there any further comments with regard to evidence?

Woman: What was the main (unintelligible) testing level (unintelligible) I seem to remember I was surprised that it was as high as it was.

(Danita): So I think in the performance gap, the - is that what you're talking about, what was the level?

Woman: Yes.

(Danita): Yes, it was pretty high. I had to keep thinking and reminding myself like low number is good, and so that was one thing I was going to ask PQA. There's so many people that could get confused with that, but that low number is good. And then this one, with the seven QHPs, they range from 57% to 84% in 2015, and then again in 2016 was just as bad as 58 to 83. And then when they did PDPs, this is going to be right now (unintelligible) for QHP, the QSR, so that looks like in the future CMS may apply it for Medicare. So they did go ahead and evaluate in PDPs with over 18 million members that again, you

know, 48% to 84% in '15 and then 48 to 80 in '16. So it just didn't improve much at all for either of those.

Woman: Thanks. Yes, that's a pretty big gap.

(Danita): And then there's actually, they do pull out a VA study and it shows on the VA that, when they actually did do the testing, they could have a reduction in opioid-related mortality by 1% for every 1% increase in the (unintelligible) screening. So that's a smaller study but that was also included in the evidence.

Harold Pincus: Any further discussion about evidence?

So I think we could move on to vote on evidence?

(Hanna): Voting is now open for measure 3541 on evidence. Options are A for high; B for moderate; C for low; and D for insufficient.

Woman: I have a question. What does it say, only eligible if (QQC) submitted, what that stands for?

Woman: That is as part of our algorithm that we use to determine whether or not what the highest rating is that can be achieved, the quality, quantity and consistency of evidence from a systematic review. And so this measure, they did provide that. So, high is the highest rating possible for evidence.

Woman: Okay, thank you.

(Hanna): Voting is now closed for evidence on measure 3541. We have 4 votes for high, 14 for moderate, 1 for low, and zero for insufficient. This measure passes on evidence.

Harold Pincus: So, now, in terms of opportunity for improvement or gap, we already have some discussion on that. Is there any additional discussion with regard to the opportunity for improvement?

(Danita): I think just add that they did go ahead and evaluate for gender, race, ethnicity and age groups. Unfortunately, in the QHP, the denominator was too small and they broke out the different groups. So they weren't able to identify a disparity. But in the Medicare population, they did do the Z test and it was found at least for gender and age there was a disparity with females and older adults are tested significantly less than their counterparts. P value of 0.0001 for both (sets).

It was thought that this probably could apply to the QHP, but it's, you know, you know, at this point, that would just be an assumption.

Man: African-Americans were tested more than Caucasians.

Harold Pincus: Other comments with regard to gaps and disparities?

I think we're ready to vote.

(Hanna): Voting is now open for performance gap on measure 3541. Options are A for high; B for moderate; C for low; and D for insufficient.

Voting is now closed for performance gap on measure 3541. We have 10 votes for high, 10 votes for moderate, zero votes for low; and zero for insufficient. This measure passes on performance gap.

Harold Pincus: Okay, now for - we'll also do a quality construct that is not - though we don't do that for this. It's on a composite measure.

So now we move to the scientific component -- reliability.

(Larry): Okay. Yes.

(Danita): Yes.

(Larry): So I can stop. I know you have a lot of questions, (Danita), that you can pick up on, because I (unintelligible) comments - in the comments section. They conducted score level reliability testing using two different data samples. They calculated across seven QHP products in 2015 and eight in 2016. And Medicare, they also computed using methods of minimum denominator (unintelligible) categories, 67 (PMPs) in 2015 and 63 in 2016.

The minimum reliability score was 0.7, and used to indicate sufficient single (string), and (unintelligible) QHPs which had at least 30 members. It was 0.85. And for PMPs, which had at least 100 members, it was 0.72. So they felt that they were both reliable.

There were some comments, and (Danita), I think you had some of them. Do you want to talk about them?

(Danita): Sure. The first I think is PQA can help me understand when defining the 90 days, looking at how CDC and others and its cumulative days of - for days' supply. However, in the definition that's used in this measure, it just says that adds up all the different days' supply together. And the comment for this is, let's say somebody has a surgery and gets seven days of a long-acting, a short-acting, and, you know, that someone also gives them some Vicodin to go with

it, as well as from (Axi), short and long, and then they got 21 days supposedly right there, they go see the doctor. And that's like, no, you know what, they need to kind of get at least another 30 days. So they end up giving another, you know, 30 days.

So, even though it's really technically 30 cumulative days' supply that they're being given, if you add up days' supply, they end up qualifying for this measure. And normally that's not how, you know, with PQA, with other measures such as the drug adherence or others, you will move overlapping days and things like that. And so I'll pause there, if you can help me understand that part.

Lisa Hines: This is Lisa Hines from PQA. Can you hear me?

(Danita): Yes, Lisa. Hi.

Lisa Hines: Hi. Thank you for your question. It's a good question.

So this measure uses a pretty standard definition of long-term opioid therapy. There's a recent systematic review published, and most definitions of long-term opioid therapy align with the chronic pain definition of pain more than three months. And most use a cumulative duration of opioid use as a criterion. So it's pretty common in the underlying research.

To your point about how measures could be - you could adjust the - or do a different methodology if prescriptions were filled on the same day, you have a really good point. I do want to highlight that this particular measure, although steward-advised PQA was developed by other measure developers, and we would, you know, like to acknowledge them, their work on this measure and impact (in HSAG), one of the things that we could consider in the (futures)

who are usual measure update process is accounting differently for prescription claims that are (filled) with the same date of service, as we do in many of our other measures. I hope that answers your question.

(Danita): It helps for future, but I don't know what that means for initial endorsements. I don't know, Peter, what does that mean for initial then, that it stays as it is?

Peter Briss: The rule is, Harold chairing this one, but the…

(Danita): Oh, sorry, Harold. Sorry.

Peter Briss: The general rule is - the general rule is you vote on the measure that you have and you can make additional recommendations toward the future.

Harold Pincus: Yes. But it is as it is.

(Danita): Okay.

Lisa Hines: And…

(Danita): Oh, go ahead. Go ahead.

Lisa Hines: And just one additional comment. Often when we evaluate the impact of (such) and changes, it's usually negligible on the actual measure rate. It's just a matter of the face validity. And so we could certainly evaluate that in the future. But again, this does align with the more common definitions in the literature.

Harold Pincus: Yes, I…

((Crosstalk))

(Danita): I know it aligns with the definition of the 90 days. I think it's the way it's going to be calculated to get to that definition is where I struggle. With all the different opioids, especially on the Medicaid population, you know, there's so much (unintelligible) on both the provider and payor, that as soon as a measure is coming out, they're getting hit pretty hard. So I just want to make sure what we are approving is really in the context of what the evidence was using to say there's evidence to support toxicity screening for long-term opioid use. So that's where my concern is coming from.

Harold Pincus: Is there any way to estimate the potential impact of this issue on the result of the measure?

(Danita): Absolutely. That is absolutely over time we refined our measures. This measure is (planned) for the quality rating system. And it needs further testing prior to implementation in other programs. So this is analysis that we could do. They have not been tested in the Medicaid population yet either.

(Jeff Gatlin): This is (Jeff Gatlin). On the other hand, the way you might look at it is that if there were two or more opioids being prescribed, that might be an indicator of poor level of care, or might indicate in fact that people are getting two or three prescribers prescribing opioids. And while I understand your point, I think you could also argue that this might actually be a better way to define the denominator. Thanks.

Man: Yes, I had the same reaction (Jeff) indicated. If after a short-term surgery, you know, people are getting three different 30-day opioid prescriptions, and this dis-incentivized that, I would consider that to be a positive outcome.

(Danita): (Unintelligible) I have thought of that too, I was just looking at the calculations.

The second one for calculation, and this was commented by someone else as well, is, you know, normally everything that's coming out is looking at it from an (MME) level. So, just make sure that we're addressing that. I guess if you do cumulative 90-day supply, it doesn't - I'm not sure if that becomes as big of an issue. But one of the commenters was, how about if it was greater than 100 (unintelligible) morphine, having (inequivalents). If PQA can just make a comment on, if (MME) calculations are taken into consideration, or thought that was needed?

Ben Shirley: Hey, this is Ben. I don't believe that we found those calculations based on the MME in the evidence. So, since we were trying to build it off of the guidelines, that wasn't included.

(Danita): Okay. And my last question, and I'm - and I can understand why it would be okay or fine, but just looking at the definition of what this is for, and it's saying that it's long-term opioid as defined as 90 days cumulative days' supply of any combination of opioid medications indicated for pain during the measurement period, and medications prescribed for MAT for opioid use disorder are excluded.

But when I look through the - at least during the validity testing as well, the HCPCs were all validated and that's all done. But what about drugs like Suboxone or methadone that could be used for both sides of detox or for pain? I didn't see any exclusions, although if (you've been) in detox, you probably want to get a urine toxicity anyways. But I just wanted to ask that for clarity just because of how it's being defined and saying that this is excluded,

however, there's drugs that could go both ways. I just wanted to just get a clarification on that.

Lisa Hines: Sure. This is Lisa. So it's based on the FDA indications, and Suboxone is not included in the measure. So, any medications that are indicated for MAT are not included in the measure.

(Danita): Okay. Thank you very much. I didn't catch. So, thanks, Lisa.

Those are my comments.

Harold Pincus: Any further comments on reliability?

Raquel Mazon Jeffers: Yes, this is Raquel, I just - I'm really struggling with the fact that the (unintelligible) is seen like such a minimum standard of quality and such an important and dangerous area. So, most clinical guidelines would recommend UDS screening pre-prescription and then every 30 days. And there doesn't seem to be a lot of thought about the potential for the (misinterpretation) of UDS screens. So there are usually like nine best practices that are recommended for reducing the abuse of opioids for pain patients. And this feels like such a bare minimum. Like, I wouldn't really give me any insurance if someone in 90 days had one UDS, that this patient was using their medication safely and there wasn't co-occurring substance use disorder that was being untreated.

Peter Briss: The only -; this is Peter ;- that might - and my counter-proposal to that is that I agree with that, although even at that, the performance is pretty low. So you might want to (start) people where they are.

Raquel Mazon Jeffers: I hear what you're saying in terms of like there's such a big performance gap anyway, but this is not for me a meaningful indicator of safe prescribing, at all.

Woman: One thing, Raquel, that you can think about it from the perspective, when I saw this measure, the first thing I asked, and I went to our - in our system and looked at the health plan, do QHP members have a copay for urine tox? Like it just brings up so many other questions that might be deterring this that just even though it's just one urine, you know, one tox screen, it could allow different benefit designs and different things to come about from this. And then maybe after the next phase, it can then say, having pre and post, versus right now having it and then there might be benefit designs that are all hampering in QHP (unintelligible) exchange (unintelligible) exchange and looking at it from - I was looking at it from that perspective, in how this could bring value for a change because of that.

Man: Well, also there's such a huge performance gap, this also raises awareness of the physician about the testing and what that means for them to review it. So I think it's a good start. I hear you that it's not the end-all, the be-all, we do much more testing than that here, but it's (unintelligible) for such a huge performance gap.

Man: I think the evidence would get thinner and thinner when you have a measure requesting multiple assessments. You know, I would be very hard-pressed to say that there's great evidence that every 30 days is better than every 60 days and 90 days. And although once a year may be too little to be ideal, I think that it gets us on a path to improvement where the benefits far outweigh the harms. Thank you.

Harold Pincus: So I think we're kind of getting away from reliability and moving into validity in a way. So, just - are there any other comments about the reliability issue?

Woman: Just the reliability of the (unintelligible) drugs (unintelligible) themselves.

Man: Uh-huh.

Harold Pincus: And do you want to say something more about that?

Woman: They're often unreliable. They can be misinterpreted. They really shouldn't be relied on as a single indicator of the presence or absence of a substance use disorder.

Harold Pincus: Does the developer want to respond to that?

Ben Shirley: Yes. I think that we would refer back to the clinical guidelines. I think that their, you know, their recommendations around the use of drug testing for people in this population are based on quite a large body of evidence. And we felt that they were quite strong.

Man: What level of - what level - this is a usability question. What level are you using this to report out at? Is this a state measure a health plan measure? I'm just thinking of rural counties where they might do point-of-care testing that might not be captured in claims data just because of feasibility of finding a lab that's close by. And I was wondering, again, what level are we intending this measure?

Ben Shirley: Specified at the health plan level.

Man: Health plan.

Ben Shirley: Yes.

Harold Pincus: Other comments on reliability?

Okay. Why don't we move towards voting on reliability?

(Hanna): Voting is now open for reliability on measure 3541. Options are A for high; B for moderate; C for low; and D for insufficient.

Voting is now closed for reliability on measure 3541. We have 1 vote for high, 16 for moderate, 3 for low, and zero for insufficient. This measure passes on reliability.

Harold Pincus: Okay. Now, comments on validity.

(Danita) or (Larry), you want to comment on that?

(Larry): Okay.

Harold Pincus: Or other members of the committee.

(Larry): (Unintelligible) with a technical expert panel consisting of three representatives from large QHP issuers and nine reps from other stakeholder groups such as measurement industry reps, clinical and non-clinical experts, patient care, patients and caregiver reps. The technical expert panel members were asked if the performance scores resulting from the measure specified could be used to distinguish good from poor plan level of quality. A hundred percent said that it could be, participated in voting and agreed that the measure was valid.

During the time that the measure was being tested, there was validity of ICD-9 (CM) to ICD-10 conversion, and we see - review the codes and agreed that the conversion to ICD-10 codes were consistent with the intent of the measure.

(Danita): Right. And I think I would just add, this is where the validity of testing of, if you did cumulative days or total days' supply, if there was a difference. But I think we've already commented on that. The validity was just testing on codes and the accuracy of that. There was no validity testing to the point of what Raquel brought up of "Does that validate if you do a urine tox screen that they truly have opioids?" And that was not performed. But for the claims that would be pulled to face validity and then the testing of the claims was done with showing that it was valid.

Woman: So, just a final comment on this. I agree that there is - it is - it has potential for - in terms of the incremental quality improvement of codes. I guess my concern is that it could offer a false sense of security or a false sense of safety, in an area where that's (unintelligible).

Harold Pincus: Other comments from the committee on validity?

(David): Hi, this is (David) (unintelligible). So the gaps of this measure is - or the evidence for this measure is driving opioid overdoses and opioid epidemic. In the Medicare population, which your data shows 79% of the Medicare patients in the study were over 65, does this measure make sense to test once a year doses greater than 90, you know, days in a population over 65? Will that reduce opioid related death in a significant way in the Medicare population? My question is whether there is, you know, that these assumptions make sense for older adults or whether we need to look at different criteria.

Ben Shirley: This is PQA. That's a good question. I would want to remind that this measure was designed and is intended for implementation in the quality rating system versus Medicare at this time. Implementation in Medicare in the future would have to be accompanied by additional testing. You know, the guidelines do apply to older adults as well, but I do want to, you know, frame the discussion around the intended implementation, which is a qualified - the qualified health systems and the quality rating system, which is a younger population.

Man: Thank you.

Jeff Susman: This is Jeff Susman. And one might argue that elders are more at risk for opioid related issues, and that often they show up reported as other conditions (unintelligible) conditions, but could be related to opioid misuse or overuse.

(David): I think that my point is, does this greater than 90-day standard apply? Would you need a different standard whether longer or shorter? So I'm satisfied with the answer for now.

Harold Pincus: Other comments?

So I guess we're ready to vote on validity.

(Hanna): Voting is now open for validity for measure 3541. Options are A for moderate; B for low; and C for insufficient.

Voting is now closed for validity for measure 3541. We have 15 votes for moderate, 3 votes for low, and 1 for insufficient. This measure passes on validity.

Harold Pincus: Now, for feasibility.

(Larry), (Danita)?

(Danita): Yes. Everything is, you know, based off of claims for the denominator qualification, as well as for the numerator qualification. So it seems highly feasible to readily gather the data.

I think except for the one point, I forgot who mentioned, of, if somebody is in a rural area and they don't have a (national) - I don't know if that becomes an electronic claimer, not if they're doing it in-house. And I think at that point we'd have to go and gather that claim in a different way.

Harold Pincus: Any other comments?

I think we're ready to vote on feasibility.

(Hanna): Voting is now open for feasibility on measure 3541. Options are A for high; B for moderate; C for low; and D for insufficient.

Voting is closed for feasibility for measure 3541. We have 9 votes for high and 10 for moderate, zero for low, and zero for insufficient. This measure passes on feasibility.

Harold Pincus: Now use and usability.

(Larry): I'm just trying to find my notes here.

Well, we (unintelligible) CMS anticipates proposing to add this measure to the QRS system in 2021, so that a scoring could be developed for 2022. And

feedbacks could provide in three different methods. And people have a chance to respond. And the technical expert panel gave feedback, which included support (unintelligible) for all individuals on long-term opioid (therapy), the inclusion of all opioids excluding those used intravenously of the epidurals.

(Danita): There was one comment by a member of the committee that said that suggesting possibly having this intended for quality improvement rather than for public reporting, because of population mix, funding mix or stratification, which could make the interpretation problematic. I'm not sure who made that comment, but if they want to expand upon it or PQA wants to provide a response of quality improvement versus going right to public reporting.

Ben Shirley: This is PQA. I would appreciate a little bit more clarification, if we could. Thank you.

Harold Pincus: Does anybody want to comment, sort of expand on that comment?

(Danita): I can't see who answered the comment, so…

(Larry): Yes, neither of us, so we don't - yes.

Harold Pincus: Okay.

(Danita): Okay.

Harold Pincus: Any other comments on use?

Mike Lardieri: Yes. Mike Lardieri. I just had a comment. Technology's changing in terms of being able to do drug screens. Now you can use swabs and those kind of things. Would they be included?

Ben Shirley: Yes. So I think in the future, as those emerging technologies are, you know, the evidence base grows and they're integrated under the guidelines, that's something we could absolutely address through our measure update panel. We do have a standard process for making those sorts of updates and maintenance to our measures as the technology has been - and the evidence evolves. So that would certainly be something that we would keep an eye on and update as needed in the future.

Mike Lardieri: Okay. If there's a new code, specifically for doing the swab, then you would just add that?

Ben Shirley: Exactly.

Mike Lardieri: Okay, thank you.

Ben Shirley: Yes.

Harold Pincus: Anything else on use?

Let's move to voting on that.

(Hanna): Voting is now open for use for measure 3541. Options are A for pass; B for no pass.

Voting is now closed for use for measure 3541. We have 19 votes for pass, zero votes for no pass. This measure passes on use.

Harold Pincus: Okay. Comments on usability?

(Danita): Yes, this is (Danita). I just had one for the, I think for PQA to just help clarify for me because (unintelligible) going to go right away for QHP. For QHP, is it on-exchange or off-exchange? Are members allowed to switch amongst plans throughout the calendar year?

And the reason I'm asking for the usability and trying to see if there's improvement, when we have this for like Medicaid or somewhere else where they might slopping around, it's difficult for a health plan to know that somebody was even on an opioid in January, February or March when they come to them in July and then are expected to try - then they're just told that there's a urine tox screening. So I just wanted to confirm for QHP, do they flip in and out of health plans throughout the calendar year?

Ben Shirley: I can confirm that specifications of the QRS do allow enrollments in different health plans, so they can switch during a measurement year.

(Danita): So then for usability, I do have a concern because health plans don't just randomly just start sending out claims to other health plans and the health plan that might get a member that didn't even known somebody was using an opioid in January, February, March, and they come to them in July, would never even known to outreach to a provider or a member that they should get a urine tox screen. So I'm not sure how this would improve the opioid concern when, if there isn't something associated with this measure to say Rx claims have to somehow be transferred. I guess I don't know how else to put that. Was that considered or discussed in the PQA group?

Lisa Hines: (Danita), thank you for raising this implementation caveat. And you raise a really good point. We, as we have other measures in the quality rating system, can create guidance around attribution, which is kind of what you're getting at, to ensure that the health plan, the patient qualifies for the denominator and, you know, they can only be numerator compliant if they (unintelligible) denominator without a health plan.

So I understand what you're getting at.

(Danita): Okay, that helps because actually in the measure when I was reading the details, it specifically said that the health plan that's a - that the measure ends with is who's going to be responsible for the measure. And that's what we concerned me, because that would mean somebody could come to my health plan in October and had all the stuff go on January through July, and I would never know, but I would be held responsible.

So, as long as that truly is the case, then I'm okay. I just didn't - it was really just called out - if it didn't say it, it wasn't even going to come to mind. But it really just said it in the measure that it's whoever the member ends with that's that health plan's responsibility.

((Crosstalk))

Man: …NQF. Just one clarifying point, the (Plus side) health plans don't experience the same amount of (churn) the Medicaid plans typically do. They do have an enrollment period. And in order for them to transfer between plans, they have to have a qualifying life event such as marriage or divorce or death or etcetera, then they have 60 days following that qualifying event to make any changes to their plan.

So we're not going to see a significant amount of changes of beneficiaries moving between plans other than that would normally be expected under those circumstances.

(Danita): Okay, thank you. And that's what I didn't know. That's what I was asking. I didn't know if QHP was like Medicaid and whether it's in or on or off exchange that people could move around.

Lisa Hines: This is Lisa. Thank you for that clarification. And also, the measure was (unintelligible) using the traditional programmatic on how the measures are handled like I said for implementation, we can modify that.

(Danita): Thank you.

Harold Pincus: Any other comments on usability?

Ready to vote on that?

(Hanna): Voting is now open for usability on measure 3541. Options are A for high; B for moderate; C for low; and D for insufficient.

Voting is now closed for usability on measure 3541. We have 2 votes for high, 15 votes for moderate, 1 for low, and zero for insufficient. This measure passes on usability.

Harold Pincus: So now we move on to vote overall regarding endorsement. Are there any final comments with regard to endorsement from any of the (unintelligible)?

I guess we can have the overall vote.

(Hanna): Voting is now open for overall suitability for endorsement for measure 3541. Options are A for yes; B for no.

Voting is now closed for overall suitability for endorsement for measure 3541. We have 18 votes for yes, zero votes for no. This measure passes.

Harold Pincus: Okay.

Peter Briss: So I think that - this is Peter back again - I think that moves us to 3492, is that right?

Woman: Yes. Thank you, Peter.

Peter Briss: All right. So I wonder if the developers would like to tee up 3492 for us please.

Ilana Richman: Sure, thank you. Good afternoon. My name is Ilana Richman. I am a general internist here at the Yale School of Medicine. I'm the measure lead for this measure here at CORE. Thank you for the opportunity to introduce the measure.

The measure under consideration today, acute care (unintelligible) opioid overdose has been developed in collaboration with the Center for Medicare and Medicaid Innovation for use in the State of Maryland. I'd like to begin with a brief overview of the - excuse me - the measure specifications and then say a bit about validity, particularly since it came up at the (SMP) meeting and since Dr. Pincus flagged it for discussion.

So this is a population health measure that captures the rate of emergency department visits for opioid overdose among individuals in a specified

geography over a one-year period. The measure uses administrative claims and it's been developed and tested in the Medicare fee-for-service population.

The denominator for the measure consists of all adult Medicare beneficiaries 18 and older who reside in measure geography, either a county or a state. The denominator is expressed in person-years, and people contribute to the denominator only during the time that they're enrolled in Medicare. And the interest is really to capture the entire population of adult beneficiaries who reside in the geography.

The numerator consists of emergency department visits for opioid overdose among - from among that population during the measurement period. Numerator then is identified from claims using a set of ICD-10 codes that indicate opioid overdose.

Now let me discuss a bit about validity which came up at the SMP meeting. So, consistent with NQF standards, we performed empirical validity testing by comparing the measure to two other indicators in the literature. The first was a measure developed by (Orc) that calculates the rate of opioid-related emergency department and hospital visits in an all-care population. And the second indicator was an age-adjusted rate of fatal opioid overdose from CDC data.

We found that in 25 states that we tested, our measure had a high correlation with both of these measures, with a coefficient of 0.74. And despite the differences in the measure definitions and the measured populations, we felt that this strong correlation suggested that we're really getting at this underlying construct of the burden of opioid overdose in a population and the burden of the opioid epidemic in a population.

During the SMP meeting, there was also a broader discussion about whether the healthcare sector can influence the risk of opioid overdose. And I think the answer to this is clearly yes. Healthcare providers and health systems play a key role in addressing the opioid epidemic. So, in what way is that the case?

First, through the use of medication assisted treatments. So I know all of you know that medication assisted treatment has been shown to reduce opioid misuse and improve retention of treatment. And it's also associated with improved all-cause mortality and overdose specific mortality among people with (self-induced) disorder.

So, on an individual level, we know that there are things we can do to reduce the risk of overdose in the clinic. On the population level, we also think that increasing the capacity to provide MAT may ultimately improve outcome. So for example, there was a recent study published in JAMA Open that found that an increase in the supply of (diprenorphine) (waivered) prescribers on the state level was associated with a reduction in opioid prescribing and an increase in buprenorphine prescription, both of which, you know, in turn are associated with reduced risk for overdose.

So that's one way in which the healthcare system can address the risk of overdose in the community.

And then the second is that healthcare providers and health systems can also play a key role by changing risky prescribing practices. So again, on the individual level there's ample evidence to suggest that certain prescription patterns, including high-dose prescribing and prescribing in conjunction with benzodiazepines is risky. And we know that it's possible to change individual physician behavior. There have been for example studies (unintelligible) changing the default settings on EMRs or providing feedback to prescribers.

So we know that we can change the behavior of prescribers, which can in turn lower risk.

And then on the population level there are things that can be done. So, for example, there was a recent study in Health Affairs that suggested that the implementation of mandatory prescription drug monitoring programs or PDMPs was associated not only with the population level reduction in opioid prescriptions but also reduction in emergency department visits for opioid-related conditions.

So we can, you know, move the needle on the prescriber level but also on the population level.

So, in summary, we think there is very good evidence that there are specific interventions that we can undertake in the healthcare system to reduce the risk of opioid overdose at the level of the individual patient, and that these changes can influence outcomes on the population level.

In closing, I'll say that I don't need to articulate I think the degree to which the opioid epidemic remains a public health crisis, and the healthcare sector is intimately connected to that crisis both in terms of its origins and its solutions. And we think this measure provides the needed tool to help (unintelligible) and maintain accountability around communities. And we look forward to the conversation this afternoon.

Peter Briss: Thanks very much. This is Peter. Actually I have a question maybe. So, on that - I'm not troubled at all by this sort of a population measure to try and trigger either population or healthcare system or even individual interventions, right? But it was interesting to me, however, that you compared this measure to at least two other sources of data that could also give you substantially

similar, it sounds like, substantially similar population level look at how communities are dealing with these problems. So, can you give us a little bit more as to what this measure adds that we didn't already have (unintelligible)?

Ilana Richman: Sure. So, why not just use that CDC data, for example, on opioid mortality? So I think, you know, there's a role for kind of a number of different kinds of measures. I think we felt the strength of this approach was several. One is that none of the other measures out there really kind of precisely captured the denominator. So, for example, in state-level mortality data and also in the (Orc) measure that I mentioned, the denominators are inferred from state population rates. Here we're really focusing on the denominator being the Medicare population, which is the population in which this measure would be used, the program in which it would be used. So, really kind of focused on the being able to measure the population accurately and then also thinking about the specific population which the measure would be used. That was I think really the driving thought behind, you know, why we need to develop a new measure versus using one of the other kind of indicators that are already out there.

Peter Briss: Thank you. And anybody else have a general opening remarks or question?

Man: This is…

((Crosstalk))

Man: …I had a quick question. Who is to be held accountable for this measure?

Ilana Richman: Yes. So it's - the planned use is in the Maryland total cost of care model. The total cost of care model holds the State of Maryland accountable for total health spending and has credits for improving quality. So the state as a whole

could be held accountable. But within that, there are groups of smaller geographies, groups of counties, that could be held accountable.

Man: And what is the mechanism for holding them accountable?

Ilana Richman: So the details of the model and how this would exactly be implemented, I'm not sure if they've been totally worked out. But basically the state can get credits for improving performance on quality, from CMS.

Jeff Susman: This is Jeff Susman, and I'll ask you now an important question to me at least, is, the intervening factors around social determinants, and if we're going to compare one county to another, or defined geographic areas to another, have you looked at the effects of varied social determinants on the rate of this measure?

Ilana Richman: Yes. So I think we can all acknowledge that the opioid crisis is not distributed evenly around the country and within states. But I will say that the root causes, the driving factors and the associated kind of characteristics, and some of these are not always the traditional social risk factors that we think about. So for example, we did look at state-level poverty rate and state-level opioid overdose rates. And the correlation coefficient was only modest. Within a state, it tends to be stronger. And actually Maryland is a good example of a wealthy state that also has very high opioid overdose rates.

So, you know, we think about as being kind of traditional social risk factors don't always (unintelligible) particularly well with risk of opioid overdose in a population.

So I think it raises an important question, is that, what exactly we would risk-adjust for and kind of the specifics of that?

To me, I think the most important thing is how the measure will be used, not so much how we thought about risk-adjusting, but how can we use the measure in a way that allows for identification of important differences and opportunities and identify those areas that are not performing well, without, you know, penalizing a geography for its history unfairly.

So I think to me the answer to that is really not so much in trying to meticulously risk-adjust for a lot of complex factors that are really hard to capture, but really thinking about how the measure will be used.

Jeff Susman: Yes. This is Jeff again. I worry, when you're looking at a state-to-state comparison that the differences in specific geographies are actually sort of washed out. Take Ohio since it's a sort of poster child for problems, and there's some locales that are handling this crisis better than others. But when you look at a statewide basis, I'm not sure what it means.

So, while I understand the rationale that you're putting forward and I think there's much merit, I also worry that, since we don't control how such measures are ultimately implemented, that there's a lot of room for (mischief).

Peter Briss: Yes. This is Peter. I actually think you relating to the last two questions, I'm a little friendlier to this kind of measure than it sounds like some of us maybe (unintelligible) primary purpose of this measure is (unintelligible) this kind of a measure is either it compared jurisdictions really or to hold people accountable in the sense that you punish them for a bad outcome. I think it's more for identifying hotspots (unintelligible) in which you might want to do more intensive intervention.

Tami Mark: Yes, this is Tami. I agree with Peter's perspective on this. It is a public - this is a public health measure, I would think about it as, it's not a clinical measure. And so we're thinking about whether the system, the whole system, meaning not just commissions but all people within the system, including public health providers, and they do (unintelligible) reduction interventions, communication campaigns, all of those resources can have an impact on this epidemic. And if we have NQF-endorsed measures, that's a tool that we can use to understand the impact of those multiple interventions on the epidemic.

And in terms of risk adjusting, I mean we know with outcome measures like this there are multiple factors that are going to affect the outcome, but we don't want to wash them out by risk-adjusting them away. I mean it's similar if we were looking at, you know, traffic action and death or death from lung cancer, why would we want to risk-adjust those? I mean, what we want to know is, what's the mortality rate? What's the adverse effect on the population?

Man: Yes. It depends on the use of the measure. If the use is for accountability or for payment, I would argue it's very important to understand those underlying drivers. If it's used in a broad quality improvement or epidemiologic sense, then it doesn't matter as much.

Tami Mark: Well, why would you want to let a county - if a county had, let's say, I mean, as the developer said, it's, in this case, it's very hard to identify a population at that particularly high risk. But let's say, I mean, the real way you want to risk-adjust it is if you're on the highway belt where all of the drugs were delivered, I mean, that's how the epidemic started in Ohio and West Virginia. So, should we risk-adjust for the fact that Ohio and West Virginia have been beyond the, you know, heroin opioid drug belt and not hold them accountable because of that? I mean, to me, that doesn't make sense, right? I mean, if you happen to

be in an area which is harder-hit for whatever reason, but does that mean that you should make the bar lower for Ohio and West Virginia than for California?

Harold Pincus: Let me ask a question of NQF staff? Is it possible, can one make a recommendation for endorsement but have a note with that vote that this is recommended as a public health measure and not necessarily for sort of a clinical measure as Tami suggested?

Tami Mark: But we don't need to do that, Harold, because it's already at the level that it's being submitted as.

Harold Pincus: Yes, it's already specified that for use at a county or a state level, so, are there any examples of sort of geographically specified measure that have been used for payments? I can't think of any.

(Sam): Hi, this is (Sam) (unintelligible) with NQF. Typically when measures are being considered for accountability programs, they'll come to our MAP process or the pre-rule-making evaluation. And at that time we'll take a look at how the measure is specified, whether it's suitable for inclusion, but of these types of payment-oriented accountability programs, I haven't seen any measures come through that have been specified at a regional level or state level that was then mis-applied. They're typically given something like additional support, pending NQF endorsement after a reclassification and testing for the appropriate level of analysis.

(Nicole): And this is (Nicole) from NQF. Just to also add that NQF's measures are endorsed at the level of analysis in which testing is provided. So in this case it would be - so the measure would be at the county and state level.

Peter Briss: This is Peter. I will want to move us forward pretty soon, unless the preliminary - yes, I've let the preliminary discussion go on for a little bit of time because I think the measure is a little unusual and I wanted us to have a chance to kick that around. Does anybody want to make additional preliminary comments before we get into the - further into the specifics? And please limit yourself to issues that haven't yet been raised.

Mike Lardieri: Yes, this is Mike Lardieri. I just had a question why just use Medicare. Is Medicare Advantage included? And if not, I think - I'm not sure that the Medicare population would necessarily be representative of a locality or a county…

Woman: Or an epidemic.

Mike Lardieri: So many people are moving to Medicare Advantage, you're not going to have a lot of people in Medicare anymore.

Woman: And I would add to that that also the overdose population is much younger. The majority of the overdose population is younger than the Medicare population.

Peter Briss: So we'll talk about (unintelligible) reliability and validity discussion. Would the developers like to just briefly respond to that issue here?

Ilana Richman: Sure. So one is that I'll say that we performed our validity testing in, you know, essentially all payor populations, and one was all-payor, one was, you know, the entire state, the CDC data. And we saw reasonable correlation suggesting that, you know, the experience in the Medicare fee-for-service population is not unique to that population but really does reflect what's going on at a broader population level.

And too, as I mentioned, you know, this measure has been developed for use in a particular program that focuses on Medicare fee-for-service, so that's very focused on that population. But we are engaging in all-payor testing, and, you know, if there were interests in extending it to a broader population (unintelligible) the testing required to, you know, expand that denominator.

Peter Briss: This is Peter again. Any other general comments or questions before we work through the specific criteria?

Man: There's only one other issue (unintelligible) discussion perhaps later on. And that's, are there intervening factors that relate to referrals or filling up the emergency room. So, perhaps you've looked at whether cities that have a very well-developed system for handling overdoses in the field or diverting overdoses to outpatient care might impact the rate of ER (use) and therefore actually give you some interesting (unintelligible) result.

((Crosstalk))

Peter Briss: Let's save that for our validity discussion, I think. And I'd like to move us to evidence at this point. So, Jeff and Tami, can you tee that up for us please?

Jeff Susman: Go ahead, Tami.

Tami Mark: Sure. Jeff, we can tag-team this. I think we had - here - a pretty good discussion. I think the developers presented a pretty good summary of the evidence. Again, for the outcome measures, what's required is that you demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. And then if you can't do that, you show variation across regions, and as the developers articulated, there's a lot

of very excellent evidence connecting health interventions to (ED) overdose death.

Peter Briss: Thanks, Tami. Jeff, anything to add to that?

Jeff Susman: I think the, again, while it might be validity discussion, it's also part of the evidence, that, are there intervening factors influencing ER utilization for opioid use?

Tami Mark: Yes.

Jeff Susman: I think that - we can defer that to - for the next discussion, but I think it is relevant probably for validity, or for the evidence.

Tami Mark: Yes. I think the question is, you know, it doesn't have to be 100% correlation. That doesn't have to be, if you do this, that will happen. I mean there will be lots of things that are going to influence the rate of ED overdose events. So that's going to be one of them potentially, but then it becomes a reliability issue, if that is such an issue that we can't reliably use the variation amongst counties or states as an indicator, so that's why it's probably good to talk about in those reliability discussions.

Peter Briss: And as it's been raised already, the developers did test comparisons with things like overdose deaths that you would think would be - should be correlated and did show correlations. And so if this were purely about - if changes in this measure were purely about differences in referral patterns or ED availability, you would expect that correlation to be (unintelligible). But…

Tami Mark: And that correlation is very, very high actually. The correlation between deaths and ED is very high. So if you look at the counties that have the high, and the states, that have the highest ED use, they're also the states that have the highest deaths.

Peter Briss: So I'd like to table that discussion for the validity discussion, and ask if anybody else has comments on evidence.

Hearing none, why don't we try to move to a vote please?

(Hanna): Voting is now open for evidence on measure 3492. Options are A for pass; B for do not pass.

Mady Chalk: This is Mady. I lost the link and I'm waiting for it to come back.

Woman: All right. Thanks, Mady.

Mady Chalk: Clicked on it again but it didn't do pretty good.

All right. Let's see if I can get it back.

Nope.

Peter Briss: And Mady, in the meantime, in the meantime, if you could still chat, that might be an option, or even if you are comfortable, a verbal vote might be an option too.

Mady Chalk: No, I'll just, I'll give you a verbal vote. Pass.

(Hanna): Thank you. Voting is now closed for evidence on measure 3492. We have 18 votes for pass, 1 vote for do not pass. This measure passes on evidence.

Peter Briss: So that moves us to evidence gap. Tami and Jeff, can you walk us through?

Tami Mark: I'll just say that the overdose variation rates varied by four times across the states from the lowest to the highest, and 10 times between the lowest and the highest counties.

Peter Briss: Jeff, anything to add?

Jeff Susman: No. I think there's at least a plausible pathway here.

Peter Briss: Anybody else has - anybody else have comments on gaps?

Hearing none, why don't we move to a vote?

(Hanna): Voting is now open for performance gap on measure 3492. Options are A for high; B for moderate; C for low; and D for insufficient.

Mady, were you able to sign back in to the voting poll?

Mady Chalk: Yes. Yes, I got back in.

(Hanna): Thank you.

Voting is now closed for performance gap on measure 3492. We have 7 votes for high, 10 votes for moderate, zero for low, and zero for insufficient. This measure passes on performance gap.

Peter Briss: So, moving to reliability. Tami or Jeff?

Tami Mark: They used a (Adams) reliability statistic among the 25 states. It ranged from 0.92 to 0.99, which is above the 0.7 threshold. And in counties, it ranged from 0.66 to 0.99, with only one state below the 0.7 threshold.

In terms of the data elements, they used the ICD-10 overdose codes, which actually are called drug (poisoning) codes. And they looked at the sensitivity and specificity relative to charts and found that they're sensitive, had high sensitivity, a little less specificity, but less specificity (unintelligible) all the stuff there. So I think there's, if anything, they're all conservative.

Jeff, do you want to add something to that?

Jeff Susman: Yes. The only thing that I did is (unintelligible) with sample reliability testing as well, and that was also high correlation at the state level. At the county level, while a little lower, it still was quite adequate. I think reliability is good.

Peter Briss: Anybody else have questions or comments?

Let's try and move into a vote please.

(Hanna): So, for this vote, just a reminder that this measure was reviewed by the Scientific Methods Panel, and so the Scientific Methods Panel did pass this measure on reliability. And so if you would like to accept the Scientific Methods Panel rating, that is the first question that we'll be voting on.

Voting is now closed. We have 18 votes for yes to accept the Scientific Methods Panel rating for reliability, and zero votes for no.

Peter Briss: That moves us to validity. We've opened this discussion already. Tami and Jeff, you want to - you want to add anything that's not been said already?

Tami Mark: I'll just say that the correlation between the measure and overdose death rates was 0.74% in the 25 states that they tested in.

Jeff Susman: Yes, I don't need to reiterate some of the concerns I hit.

Peter Briss: Okay, thank you. And anybody else have - anybody else have additional comments or concerns that haven't already been raised?

Let's try moving to voting…

Raquel Mazon Jeffers: I'm so sorry. This is Raquel. You said - the measure developer said that you would be willing to consider expanding beyond the Medicare population? What would be required to do that? Is that just for our future or is that something you could - I don't know how operationally that works now.

Ilana Richman: This is the developer. Oh, go ahead.

Peter Briss: Yes. So this is Peter. So again, the general rule is we have to - we have to vote the measure that's in front of us. So the testing data that's in front of us is the testing data that's in front of us. You know, the developer has, I've heard the developer express a willingness to do broader testing going forward if that's stuff that people would like to recommend. And so we could recommend as a committee as a going forward issue. It shouldn't affect the voting today.

Would either the developer or the staff like to add something to that that I didn't already say?

Ilana Richman: Sure. Yes, meant it as a nod to the idea that testing the broader population makes a lot of sense. But the measure is developed with a specific population in mind at this point.

Peter Briss: And there were - and again, there were correlations with other broader populations, is that correct?

Ilana Richman: Absolutely. Absolutely. Both of the - both of the indicators that we looked at for empirical validity testing used essentially an all-payor, an all-population population.

Peter Briss: Okay. So, any other questions, comments or concerns that haven't already been raised?

Let's move to voting please.

(Hanna): As a reminder, before we do move to voting, the Scientific Methods Panel did look at this measure. The validity rating from the Scientific Methods Panel was low. And so the first question we are asking is whether or not you accept the Scientific Methods Panel rating for validity. Again the vote - rating for validity was low, from the Scientific Methods Panel.

Woman: I mean, I know we're running out of time, but I really think we need to revisit the role of the Scientific Methods Panel, because I really don't think it was useful for this.

Peter Briss: Yes. So I wanted to say that too. So, staff should correct me if this is wrong, but if we were to accept the Scientific Methods Panel's rating which is low, that would mean that the measure would fail on that must-pass criterion and we would end our - we would not pass this measure, it would not pass, it would not collect $200.

So if you vote - if you - essentially, this is a little confusing, but if you accept the panel's rating and you vote yes, that means the measure fails at this point, I think. If you vote no, then we do our own reassessment. Do I have that - is that right? Staff?

Woman: Peter, that's correct.

Peter Briss: Did everybody understand…

Ilana Richman: I just want to add, from the measure developer, just a reminder that the Scientific Methods Panel did not have much of the information about the evidence that's in front of you at the time that they were evaluating the measure.

Woman: And they're also not subject matter experts, which is why you have a behavioral health committee doing this. So I think it's - we really need to revisit that process.

Peter Briss: So in this - on this vote, if you vote no, that allows us to have our own vote on validity. If you vote yes, you agree that there are deal-breaking validity concerns and the measure fails. Is everybody clear about that?

Man: Thank you.

Peter Briss: So let's vote please.

(Hanna): Voting is now closed to accept the Scientific Methods Panel rating for validity. We have 8 votes for yes, 11 votes for no. This is consensus not reached, so we will move to a re-voting on…

((Crosstalk))

Woman: So we're going to move to the next vote, so we're going to revote on validity. And the possible ratings will be high, moderate, low, or insufficient.

Man: And so, just remind us, if we were to vote low, and that was - or insufficient, those would be suggesting the scientific acceptability, validity was not enough, correct?

Peter Briss: That's correct. Either of the - either of those means that the - if enough people vote for either of those, then the measure would fail on the validity criterion.

Man: Okay, yes. Thank you.

Woman: And validity in this case means what?

Peter Briss: It measures what it purports to measure.

Woman: Meaning that it meant it purports measure overdose, ED overdose rates.

Peter Briss: Yes. It's essentially, does it validly measure geographic areas with a higher or lower rate of overdose, essentially.

Woman: Of ED visits for overdose. So, does it validly measure the variation across geography and ED overdose, yes?

Peter Briss: That's correct.

Woman: Or visits. Yes.

Peter Briss: Yes, that's essentially correct.

Woman: Okay. We're not voting on whether it's reflecting the quality of the medical care in that region, to…

Peter Briss: No.

Woman: …be clear on that. Okay.

Woman: I think we're looking for one or two more votes. Has everyone cast their vote at this time?

(Hanna): Voting is now closed for validity on measure 3492. We have 2 votes for high, 9 votes for moderate, 6 votes for low, and 2 votes for insufficient. This measure is consensus not reached.

Woman 1: Okay. So the results of the vote for validity, as (Hanna) mentioned, are consensus not reached. So in this case, we will continue moving forward discussing the rest of the criteria starting with feasibility during our next call. We will not do that overall suitability for endorsement vote, even if the rest of the criterion passed. And that's where we will pick up for our next meeting.

We very much appreciate everyone's time and the extra few minutes you've spent after the hour. We really appreciate it.

Harold or Peter, any last-minute remarks for the group?

Peter Briss: Do we need to open for public comment before we adjourn?

Woman 1: We do. Thank you. So if there are any members of the public on the line that would make - like to make a comment, please feel free to (do so at this) time. Okay. Hearing no comments and we have received nothing in our chat box, I think we can move to conclude our meeting for today. Thank you everyone for joining us. And we look forward to continuing our conversation on February 5th.

Peter Briss: And this is Peter. Thanks everybody. This was - we got a lot of stuff done today, I appreciate everybody's terrific attention and collegial discussion. Thanks.

Harold Pincus: And I agree with Peter, as usual.

Man: Thank you, peerless leaders.

((Crosstalk))

Woman: …leadership. Yes. Thank you.

Man: Thanks guys.

Woman: Bye-bye.

Woman: Thank you. Bye.

Woman: Thanks. Bye.

Man: Bye-bye.

Woman: Bye.

END