NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
June 26, 2019
11:38 am CT

Man: Breaker 1-9, Breaker 1-9, this is Hardhat.

Sheila Crawford: Good morning everyone. To the standing committee members, if you can while you're waiting for the meeting to get started, go into your Poll Everywhere Link just to ensure that you have access. The link is in the email. I just sent the email out again right now, otherwise it's in your meeting appointment or in the email from yesterday from our Project Team. And if you're having any issues, just let us know.

Operator: The conference has been muted. The conference has been muted.

Sheila Crawford: All right. Good morning everyone. Welcome to the Primary Care and Chronic Illness Standing Committee meeting for this Spring Cycle. We're so excited to see everyone in-person, old faces and new faces. I'm going to go ahead and pass it to our Co-chairs, (Gail) and Adam to kick off the meeting.

People on the phone, just a reminder if you're not speaking, please put yourself on mute. Thank you.

Man: So, thank you (Unintelligible). Good morning everyone. It's really a pleasure to get to see everybody and work with you again. I want to thank all of you for all the work that you've done. I actually spent a lot of time yesterday rereading all the comments that I know many of you put into the Measure Evaluation, so I really appreciate all the hard work.

We have a very full agenda today. Tough to get through it all, so we'll quickly get started here pretty quickly. I'll turn it over to Adam.

Adam Thompson: I would just echo the welcome and echo we have a lot of work to do. So, with that, I will pass it to NQF.

Sheila Crawford: All right. We're going to go ahead and pass it to our Senior Director, Sam, just to give a brief introduction of the day and the (steps).

Sam Stolpe: Good morning everybody. So, you have the agenda in front of you. As you can see, and as you already know, we do have an ambitious schedule. We're going to get through a total of 10 Measures today and a big thanks to our Measure Developers for coming in to assist us as we evaluate these.

I wanted to just briefly introduce the staff. I'm Sam Stolpe. I'm the Senior Director for this project. We have (Suzanne Tebar), she is our Senior Project Manager; (Harol Duvualle), who's a Project Manager that you just heard from; as well as (Asaba Denwulquafor). I'm sorry, I really struggle with your name all the time and one of these days I'm going to get it right. So, thanks (Asaba) for your patience with me.

We also have our Senior Vice President, Elisa Munthali, who is going to do our Disclosures of Interest.

Elisa Munthali: Thanks, Sam. And I wanted to welcome you and thank you for being on the committee. And so, what we're going to do today is to combine our introductions with the Disclosures of Interest. And you probably remember when you were named to this committee, we asked you a number of questions about relative, relevant information that was related to the work that we're doing today and so what we're asking you to do for this meeting, is to orally disclose what you provided to us in that form.

Just a couple reminders before we go around the room and I understand we have one committee member on the phone, we are interested, not just in paid activities as they're relevant to the work in front of you, but also those activities that are unpaid. We also want you to know that you sit on this committee as an individual. You do not represent the interest of anyone who may have nominated you for the committee or your employer.

And probably the most important reminder is, just because you disclose does not mean you have a Conflict of Interest. We go through this process in the interest of transparency and openness. And so, I'll start with your Co-chairs. I'll first start with Adam and then we'll go around the room clockwise and then I'll call on the committee member that's on the phone after we do disclosures in the room. So, Adam.

Adam Thompson: Good morning. Adam Thompson. No disclosures.

Dale Bratzler: Dale Bratzler, University of Oklahoma. No conflicts or disclosures.

Elisa Munthali: Thank you very much. And, before we start, I forgot to say, when you introduce yourself, let us know who you're with and tell us if you have any disclosures. So, Katherine.

Katherine Gray: Am I on? Katherine Gray. I'm with Sage Health Management Solutions and we provide clinical decision support in the area of imaging and a few other things. And I don't have any specific Conflicts of Interest other than working for an organization that does the clinical decision support development.

Catherine MacLean: I'm Catherine MacLean. I'm the Chief Value Medical Officer at the Hospital for Special Surgery in New York. And I'm also a long-standing member of the American College of Physicians Performance Measures Committee.

James Rosenzweig: I'm James Rosenzweig and I don't know whether I disclosed this, but I am the Chair of the Quality Improvement Subcommittee of the Endocrine Society and in addition, we are now working with the Endocrine Society to Develop Performance Measures for Hypoglycemia. In addition, I did some work many years ago with PCPI. And more recently I was on an Advisory Committee for Measures for the Prevention of Diabetes.

Faith Green: Hi, good morning. My name is Faith Green. I work with Humana focusing on Clinical Quality and I do not have any disclosures.

John Ventura: John Ventura. I practiced chiropractic for 30 years, now own a consulting company. And I hold a faculty appointment with the University of Rochester's Department of Family Medicine at the School of Medicine. And I have no disclosures.

Don Goldmann: Hi, Don Goldmann from the Institute for Healthcare Improvement in Boston's Children's Hospital. I don't think anything to disclose, but since it's related, I'm the Chair of the AHRQ National Advisory Council.

Rishi Singh: Hi. I'm Rishi Singh, Cleveland Clinic in Ohio…

(Gail): Microphone, please.

Rishi Singh: Sorry. My name is Rishi Singh from Cleveland Clinic in Cleveland, Ohio. I'm on the Board of the American Society of Retina Specialists. My disclosures are, I do numerous clinical trials, none related to any of the measures we're discussing today. And my financial disclosures are, again, for consulting which are not related to any of the discussion points we have for today's measures.

Scott Freeman: Good morning, Scott Freeman, American Academy of Ophthalmology. No relevant financial disclosures.

Bill Curry: Hi, I'm Bill Curry with the Department of Family Community Medicine for Penn State University. I'm Vice-chair for Population Health and we study the use and put quality improvement measures into effect based on some of these measures. I have nothing else to disclose.

Vicky Shanmugam: Good morning. Vicky Shanmugam, I'm the Chief of Rheumatology at the George Washington University here in town. Welcome everyone to Washington, D.C. My disclosures actually pertain to the American College of Rheumatology. I am the Chair of the Annual Meeting Planning Committee for the American College of Rheumatology and I also serve on the Committee for Education.

(Starlin Hagengrading): (Unintelligible).

(Gail): No. The right button.

(Starlin Hagengrading): That's what I pressed. Oh. Technology. I'm (Starlin Hagengrading). I'm a Pharmacoepidemiologist from Illinois. I'm an Independent Consultant. I am on contract with the Illinois Department of Public Aide and the Illinois Department of Public Health to review their Quality Processes in that state. I also need to disclose that I am an appointment member for Pharmacy Quality Alliance. I have not worked on any of the measures that we're reviewing today. I am a specialist and they bring me in on work groups and reviewing those areas in endocrinology, cardiovascular, and other chronic health diseases. Thank you.

Woman: Kim Elliott and I reported measures for health plans for about seven, eight years and I was responsible at the Medicaid Program in Arizona for determining which measures we would implement across all of our programs. And now I work for an External Quality Review Organization for the past three years, where I actually validate performance measures for (UREC), NCQA, for (BATES). And also work with state implementing performance measures.

Jeff Lewis: Hi, my name is Jeff Lewis, (LEO) Community Health and I have no conflicts or disclosures. Thank you.

(Gail): The right button.

(Unintelligible Levine): Hi, I'm (Unintelligible Levine). I'm from (Unintelligible) Health in Long Island, New York. I have no conflict of interest. I run the Quality for Rheumatology Division (unintelligible) our facility.

(Lindsey Bachford): Hi, good morning. I'm (Lindsey Bachford). I'm a family doctor from Houston, Texas with Memorial Hermann Healthcare System, well, for a few

more days and will be transitioning to work for IORA Primary Care next week. No disclosures.

Daniel Greninger: Hi, I'm Daniel Greninger. I work for Kaiser Permanente. I'm an Ophthalmologist. No disclosures.

Elisa Munthali: Thank you very much. And I understand William Taylor, are you on the phone?

William Taylor: Yes, I am. So, I'm William…

Elisa Munthali: Hi.

William Taylor: Hi. Do you want to hear my disclosures or identification?

Elisa Munthali: Yes, please go ahead.

William Taylor: Great. So, I'm the Medical Director for Medical Education at Atrius Health. It used to be Harvard Vanguard Medical Associates. And I'm also connected with Harvard Medical School where I'm an Associate Professor of Population Medicine and Associate Professor of Medicine. And at (Bethus) (Unintelligible) (Guinness) Medical Center at Brigham and Women's Hospital and I have no disclosures.

Elisa Munthali: Thank you very much. Before I turn the meeting over to Suzanne, I just wanted to remind you that if at any time you remember you have a conflict, we want you to speak up. You can do so in real time or you can approach any one of us on the NQF Staff or anyone of your Co-chairs.

Likewise, if you believe that anyone on the committee is acting in a biased manner, we want you to speak up. So, thank you.

(Suzanne Tebar): Okay. Thanks everyone. Before I talk to through the (unintelligible), just a quick reminder, we do ask you to use your microphone so that it picks up for the people on the phone, as well as the recording. So, and then please turn it off when you're done so, we can only have three live mics at once.

So, I'm just going to - I also want to quickly acknowledge and thank the expert reviewer pool for this committee. As you know, we pull folks in every cycle depending on what we have on the review. So, we do have many people on the committee who are not active this cycle, and we just want to thank them and acknowledge their role in our work as well.

So, I'm going to briefly talk about the evaluation process. Just a reminder, I know you mostly have all been through this before. We will start off each measure with a very brief introduction from our Measure Developer colleagues. They will introduce the measure and address anything that we ask them to address based on your comments that were submitted in the pre-evaluation review.

Then we will ask the Lead Discussant to introduce each measure, talk about the issues that were raised by the staff and the committee in their review and then open it up to the other discussants in the - for each measure and then to the full committee for each of the criteria.

As you know, your role during this meeting is to act as a proxy for the NQF Membership and evaluate each measure against each of the criteria and then make a recommendation to the NQF Membership, the Public, and the Board

whether each measure should be endorsed or not. We also ask you to oversee the General Portfolio of Primary Care and Chronic Illness Measures.

Just general ground rules. We ask that you have reviewed everything by now. That you look to our criteria to make your recommendations. And that you tend to keep your comments concise and nonduplicative and just try to make sure that we get the full range of opinions, but not too much duplication.

I'm going to just remind you briefly about what the NQF criteria are and what the differences are for New Measures versus Maintenance Measures, those that are coming back for re-review. The important criteria looks at both the evidence, the quality, quantity, and consistency, or an established link with a process in the healthcare system for outcome measures.

And Gap looks at how much opportunity for improvement remains, what the performance generally is, or whether there are gaps in some populations, be that demographic groups, geographic regions, types of insurance, anything like that.

For Maintenance Measures, we ask that either developers say that the evidence hasn't changed or that it has. If it has changed give us that update and for - but we can decrease the emphasis on that for maintenance measures versus new measures.

And then for Gap, we place a greater emphasis on that. We want to see that there is still room to improve for our Maintenance Measures versus our New Measures.

Our Scientific Acceptability criteria are reliability and validity. We want to make sure that the specifications are fully updated for Maintenance Measures

and that they are precise for all measures. And we now ask that Measures look at the risk - for Social Risk Factors and the Risk Adjustment. And I'm going to pause here and turn it over to Sam for a couple of remarks on our Measures that we're looking at today.

Sam Stolpe: Thanks very much, (Suzanne). So, there's a recurrent theme, that, undoubtedly many of you have noticed inside of the Reliability Section of Staff Reviews for these measures. The thing that has come up is that we have a (criteria) around Level of Analysis for separation when a measure has been specified for more than one Level of Analysis. The expectation is that the testing reflects that. So, those need to be separate.

This isn't a new (criteria). This is something that's been around for a while. But we haven't always been as stringent in requesting the Measure Developers ensure that their analysis are not presented in aggregate. So, NQF is becoming more insistent on that. And we're asking our committees to do the same.

Now this isn't guidance specifically from staff either. This has come as a clear directive both from CSAC, as well as from our Scientific Methods Panelists who have identified this as a potential problem inside of a liability. It becomes an issue, in particular, when the analysis is specified at the Clinician individual and Clinician Group Practice level.

The reason we don't want those in aggregate is because Reliability, especially Score Level Reliability, ends up being a function of the number of events that occur. And so, the issue becomes particularly poignant when there's few events and where a concern is. That if you're an Individual Provider, then those are more likely to have fewer events than our Group Practice Level Provider setting.

And given that's the case, we need to see those separated out so that we can have the confidence, especially for payment purposes, when there's a lot on the line that the reliability, when we say that one provider is performing better than another, that we can say that with the kind of confidence that we would say, we would pay this provider more. It can't just be left up to some stochastic process.

So, that's why we're being more insistent on that and NQF is sticking to our guns. Now that being said, because this may appear like splitting hairs, if they are in aggregate and they can demonstrate a certain level of confidence and reliability and provider ranking, then we're okay with allowing a group level specificity.

So, although all of these is marked insufficient, it's marked insufficient because those analysis are not separated. But if the committee wishes to move forward with allowing a group level analysis, that would be acceptable by NQF's criteria.

Thirty questions on that. I just, I realized that I'd talk for a minute. Okay?

(Suzanne Tebar): Okay. Thanks, Sam. The remaining criteria are Feasibility. How easy is a measure to implement? What are the costs associated with it? We have a separate Feasibility Assessment for E-measures and we are going to be looking at some E-measures today.

We have NQF's E-measure expert will be joining us to assist if we have any questions there. And then Usability and Use. We look at how measures are being used and how measures are taking into account feedback from users

from those being measured, as well as any unintended consequences of the measure.

For both New and Maintenance Measures, we have no difference on Feasibility, but for Maintenance Measures, we do look for measures to be actually in use after endorsement. And so, we have an increased focus on that. Any questions on our criteria? Great.

I'm going to talk briefly about voting. So, as you know, the way the process works is we discuss each of the criteria and then we vote. And then we continue with it to the next criteria. So, we start with Evidence and discuss vote, then move onto Gap, discuss vote and so forth.

Importance, so both Evidence and Gap are Must Pass. Scientific Acceptability is also Must Pass. And Use is Must Pass for Maintenance Measures. If a measure does not pass one of those Must Pass criteria, then we stop and discussion does not continue.

What Pass means is that a quorum of the committee, greater than 60% of the committee votes either Pass, or High, plus Moderate. So, a quorum is 66% of the committee. We have everybody here, almost, so we are all set with quorum. That's 14 people for the committee. So, actually, if you have to leave early, please do let us know at a break so that we know if we might lose quorum at the end of the day.

As I said, Pass or Recommend is greater than 60%. Consensus not reached is everything between 40 and 60% inclusive of both 40 and 60. And Does Not Pass is less than 40% Yes votes.

Measures that don't reach consensus continue through the criteria discussion. The committee does not vote on an overall recommendation for endorsement. You will reconsider at the post-comment call in September after getting additional input from the Developer on any questions that you have, as well as feedback from the NQF Membership and the Public.

So, I'm just going to briefly, we've got 10 measures under review, as has been said. Here's the List, I know you've all looked at them and that you are familiar with them. And we - so we have a full agenda and then just briefly, we did have one measure that we're looking at today went through the Scientific Methods Panel. It is a complex measure and the Scientific Methods Panel did agree. It passed all Reliability and Validity criteria.

You will have a different option when you vote on this measure. You can vote to either accept the Methods Panel Recommendation or you can vote to make you own recommendation. And then you would vote whether it's High, Moderate, Low, or Insufficient.

And then the Methods Panel did also look at another measure, but it didn't pass the - it didn't reach consensus. And, actually that measure has been deferred. So, you'll look at it in a future cycle.

With that, I will pause and see if there's any question before I turn it over to our Developer for the first Measuring Introduction. Any questions? Okay. So, I think we can proceed to Measure Evaluation.

Lisa Suter: Great. Thank you so much. My name is Lisa Suter. I'm an Internist and a Rheumatologist at Yale and I'm here representing the ACR for our three measures. They're part of a larger suite of Rheumatology Measures that are intended to assess Best Practices for Rheumatologic Care. And then we are

working with ACR - sorry with the NQF Incubator Project to move those towards more specific process measures that actually define, you know, what do these activity assessments and what functional set assessment tools we're collecting on patients and then move those into outcome measures.

So, these are really foundational, but kind of initial quality measures, that you're coming back to again today. And I want to thank the committee for their time today and the opportunity to speak on your very detailed review of our measures.

I'm going to address two, I think, overarching themes on the measures in my introduction and then there are a lot of more specific questions that I'm hoping I can address as they come up from the committee during discussion. Those two are the concept that these were submitted as eCQMs, but moved into sort of an E-measure category.

And the second was, as Sam raised, Individual Clinician Level Reliability testing. So, these measures are part of the RISE Registry, which represents about 30% of Rheumatologists in the nation, which is the largest Rheumatology Registry in the Country. And that's prior to negotiating with EPIC to allow coordination with that EHR System.

We represent primary care providers - excuse me, Rheumatologists with over 30 different electronic health record types. And the way we have built our Measures is to highlight functionality and clinical workflow for the individual clinician. Which means that we have not imposed standardized data elements on them.

We have worked with each individual clinician. There is actually a handholding onboarding process with every single practice, to map the

elements to their respective EHRs, their respective clinical workflow, and their notation system. So, that means, if they have standardized lab values, we're pulling standardized lab values that's pretty consistent.

But if they're documenting their PPD testing, or they tuberculous testing in their notes, we work with them to do an iterative validation process with them individually to map that accurately. So, this is a rigorous process. We do it for free for our members because we think that Quality Measurement is that important and Feedback is that important to our Rheumatologists.

It also means though, that as CMS has defined and now NQF has defined, eCQMs as, you know, (unintelligible), closed, specified (unintelligible) tested measures, these measures don't meet those criteria. They have, however, been tested in the way that the electronic data has been compared to chart review by humans, by clinicians, to ensure that each data element is valid and that when we think we're pulling information from the medical record, through our electronic process, we've validated it to make sure that it is, in fact, accurate.

So, just a reflection that these don't meet the technical eCQM criteria, we are trying to move towards that because we know that that's the future. But we started from a position of getting practice's information that's useable in the moment for them and not upending how they document all of their clinical encounters. And we've had a lot of positive feedback from our members about that.

The Registry provides a Dashboard to all participants. They can drill down to the patient level. They can separate out practice information at the end of the drill level. There's a lot of information there and we are revising it on an ongoing basis with input from our members.

And then the second issue is Individual Level Reliability testing. So, not fully appreciating the distinction from the testing expectations, because last time this came through, we submitted reliability data at the practice level and we supply practices with practice level information.

We were able to, in the short period of time between when we heard this feedback and today, do the testing at the Individual Level. We supplied NQF with a Memo yesterday about that. The results are essentially identical as long as you have a volume cutoff of 10 patients for each of the three measures, which is a pretty low volume cutoff for reporting. And all of the, essentially, at volume cutoff of 10 patients, every practitioner has a Reliability of .7 or greater.

I'll pause, I guess I'll hand back over to the committee and answer questions as they come up. I know there were a lot of specific questions. I'm happy to address those as they come.

Man: So, just a side note. We do have a copy of that available for the committee, which we can bring up that shows that analysis and we'll (review) that for the reliability discussion.

Adam Thompson: Thank you so much. So, with that, we'll begin by asking our Lead Discussant, Katherine to briefly summarize the measure and then we'll move into a specific conversation around Evidence.

Katherine Gray: Yes, the particular Measure that I'm the Lead Discussant for is 2522, which is Rheumatoid Arthritis, Tuberculosis Screening. And the level of analysis is by clinician, group practice, and individual. So, two different clinician groups.

And the, I'm sort of torn here, because we have, you know, if I calculated this right, I had about 30 minutes to do all this stuff. So, I'm going to kind of just go to the heart of everything. If that is not good enough and the group wants to have more dialogue, I'd be glad to do that too. But I'm just going to try to skip to the - if I look at from the evidence, it was judged as Moderate by the Panel - excuse me, by the staff - what should we call the review panel before, was that the staff or is that the Scientific Committee or what is it called? It's the staff, okay.

So, it was judged as Moderate, which is, you know, pretty good. But one, the comments that came back from the other committee members, or the discussants for this particular panel, there were a couple of things that I thought were really important. So, this is kind of how I'm doing this. If you don't like it, speak up, so.

Okay. And what was of concern in the comments is that the technique of using the RISE Registry and basically, these are my words, but biased, in that it represents smaller practices and not larger practices. So, that seemed to be a concern to me from my background in measurement.

So, that's kind of one of the issues that people had raised. So, I'm just going to bring that up. Does anybody want to talk about that or not?

Man: I'll bring it up. It's the same issue that will come up with the next measure that we're going to look at and I - the way I kind of rationalized it - and the Developer can help with this is - it does represent 30% of practicing Rheumatologists which I think is pretty substantial.

And secondarily they did work with over 30 EHR vendors. So, I would think that they've covered a substantial number of those out there, when you

consider that the majority of practices are within about five or six different EHRs.

Vicky Shanmugam: I'm going to comment on that actually. I (shouldn't make probably) its predominantly private practice Rheumatologists, correct?

Lisa Suter: At this time, it is, but we have finally negotiated with EPIC and we have an Individual Academic Center who led that charge. It's been a very long negotiation to get EPIC to navigate this kind of data exchange. And since the predominant number of academic centers are EPIC, we anticipate that there'll be a lot larger number of academic centers participating.

In terms of this measure, the mean number of patients per practice of rheumatoid arthritis patients qualifying for this measure is over 200. It's 208. So, I would argue that these are not all, just small practices. They are a range, but they are predominantly private practice. The range of volumes ranges from one to almost 15-hundred. So, there's a huge range of practice size.

I think we represent geographically, demographically, a really diverse set of practices. And, while you might think that the practices that are most likely to use Quality Measures are the ones that are going to adopt the registry, I think actually the payment legislation of MACRA has really pushed and we've really had an explosive growth in the last couple of years and so we have a huge range of practices participating in RISE and not all of them are really engaged in Quality, so we've had a great opportunity to engage and educate our constituency through RISE.

Man: I would like to comment that there are other large EMR vendors and not all of them are so willing to put measures like this across their whole platform. Also, working in an academic medical center, our EMR, although it's supplied

by one of the major vendors, it has been modified so many ways and by so many groups asking for special things, that any upgrade is a tremendous effort and difficult to do.

So, I think that there will be challenges in the years ahead to bring the work that the RISE group has done, which is, I think, laudable. But I think there's going to be large numbers of rheumatologists that will struggle with getting their EMRs into a position to do what they're doing with EPIC.

Man: Yes, hi. So, first I apologize for not having followed the proper procedure to submit comments in advance, which I somehow forgot that I should do. But I've made many comments on my (sheet) here. But I have an overarching question around the use of the Registry, which may seem a little philosophical, but it's important to me.

I'm all for using Registries as a reliable source of data for improvement, but in my experience, registry data is kept within the community of the practicing physicians in that registry. So, what I'd like to know is, are these registry data on performance going to be publicly available to patients? What about the 70% of practicing, or whatever it is, rheumatologists who are not part of the registry? What is the organization doing to encourage, monitor, ensure testing for those patients?

I mean, this is something that's on every TV ad for these drugs in America and it seems to me that there is not just a rheumatology and registry interest here, but a population health issue. So, I wonder what the comment would be on that?

Lisa Suter: These are all great points and so first of all, this Measure is - we've been in conversations with CMS. They asked us to combine it with another screening

measure for these kinds of medications. But we're in negotiations with them about the fact that there're different denominators across those screening measures and so we are optimistic that's going to be incorporated into MIPS. And once it's incorporated into MIPS there will be public reporting for practitioners submitting this measure.

I think the broader issue about flexibility and nimbleness with EHRs is an important one. We've started from the perspective of really trying to work with our practices. We are moving towards recognizing the need for standardized data collection and standardized fields and so we're working - we already have created recommended standard practices for our membership about what data elements to collect and how. And so, I think we're hopeful that we can move towards a much more traditional eCQM in the future.

In the meantime, we're working with all EMRs. We have push and pull data processes so that we can push and pull data depending on what is acceptable to the particular EMR vendor. You know, I think, it's an important part. We're working with CMS to think about how we can make more rheumatology measures public and feasible and this is something that CMS is struggling with.

We're probably not the only registry, you know, our vendor that helps support our registry supports, I don't know, something like 20 other society registries, not all of which are as extensive or longstanding as ours. And they're all struggling with this. You know, how do we balance the burden on the practices with the interoperability.

So, we appreciate that. I think in terms of your direct question, yes, this particular measure will be public and we're working on all fronts for both sides.

Man: Follow up really quickly. So, I'm assuming that many rheumatologists are members of the society but not in the registry, is that correct?

Lisa Suter: Yes.

Man: So, what is the society doing to - apart from this measure, just from an ethical and health perspective? What is the society doing to encourage the use of tuberculosis screening in those who are not in the registry, but are your members?

Lisa Suter: So, we - first of all, I should have acknowledged that on the phone, we have Amy Turner and (Tracy Johansen) who are HR staff who work with the Quality Measures Subcommittee, of which I am Co-chair and support measure development and quality efforts across the college.

So, there are a lot of opportunities and ways in which we support clinicians. We have Guideline Projects. We have Best Practices. We are at the (unintelligible) registries at every meeting that the ACR holds. Both its annual meeting and all the smaller scientific meetings. We're working with practices, even who aren't participating in the registry, to collect standardized data and thinking about data quality.

So, it's a long list, you know, in terms of newsletters and email blasts and things like that. There's a whole educational arm of the ACR that is focused on Quality Improvement both in the context of MACRA and outside of it.

I don't know if Amy or Tracy, you want to add anything?

(Tracy Johansen): This is (Tracy Johansen). I would just add that the Tuberculosis Screening Measure is - we have essentially a sister version of it that we are steward for in the MIPS Program. So, that measure - we maintain both MIPS and the QPP version of this measure and work to make sure they're harmonized and the QPP version of the measure is also available to - excuse me - rheumatologists that - or rheumatology providers that would like to submit this measure as part of their MIPS submission, as well and track it within MIPS as well.

It can be difficult when you're trying to do - work on measures like that outside of the registry, because the HR systems are not so friendly to rheumatologists who want a report on measures outside of the basic standard of care. So, as Lisa mentioned, we are still working hard to get rheumatology providers to pay attention to rheumatology specific quality measures and the quality of care guidelines that have been set out by our many volunteers and experts and - excuse me - and get as many rheumatologists and rheumatology providers connected to RISE with the help of greater access to these rheumatology specific measures. They may not have as easy access outside of our registry.

Adam Thompson: So, just a real quick process point. As folks have comments to make, if you could be sure to raise your name tag for us so that we can kind of see and be able to, sort of navigate the conversation. And with that, I would ask, if we have any of our other assigned discussants or any other members of the committee who have anything additional that they would like to add to the discussion around Evidence?

Woman: I just wanted to make a comment that somebody made across this table that at an institutional level, I think we've run across some hurdles in terms of trying to integrate data across a generalized registry through the ACR. And my

impression and the question to you is really, is that data available real time to people who are not going to be participating in the RISE Registry? Because as we are trying to incorporate these measures in our practice, trying to get some realistic expectations of what to focus on for our metrics, we have no guidance if we don't have that data available right now.

Lisa Suter: I'm - let me just see if I'm understanding your question. You're asking, do people who have - who are not participating in the Registry have access to real time Quality Measure Results? So, I think that's a question for each individual practitioner. They do in RISE, right? Their Dashboard is there and gets updated and they - that information is fed back to them regularly.

Woman: My question was for people who are not in RISE. So, if an institution is not part of RISE Registry, but is implementing these Metrics, what are their goals? What standards do they adhere to?

Lisa Suter: So, I think, in answer to your question is, it's hard for individuals to - many practices do adhere to these Quality Measures on their own and they submit them in different ways. So, some of these measures were in PQRS and were submitted in claims code and so practices have ways of identifying for themselves, whether they're screening for tuberculosis.

You know, one of my Rheumatology Clinics is in the VA and we have a process for measuring that. Whether or not you have that data available to you at the point of care is entirely dependent on your institution and your ability of your institution to work with you to make real time data available. It has very little to do with the Quality Measure itself. It has to do with IT expectations within the - and flexibility within the EHR. I may not be answering your question and I apologize if that's the case.

Adam Thompson: So, very quickly, I'm sorry, we're kind of getting into the Feasibility discussion. Which we will get there, don't worry. But looking at the time, I want to make sure that we sort of go step-by-step through this, or else the whole day will be chaos.

So, what I would ask is that if folks have comments that specifically relate to Evidence at this time, you can keep your name tags up. And if not, if you have additional elements that we're going to discuss later, just hold those comments until we get to that section.

Vicky Shanmugam: So, the only comment I was going to make was to refocus the discussion. But our job is not to assess whether or not, at this point, people can measure it, but more to assess whether or not there's evidence to support (sustain) or Quality Measure? Which I thought that there was for this specific measure.

Adam Thompson: Thank you.

Man: I'm just curious for future education in my participation. I would have thought this was a slam dunk on Evidence. What do we - do we need an RCT to show that you can get TB when you're on these drugs and that if you screen, you can treat and prevent TB? What was - what held the staff back?

Adam Thompson: So, the staff actually rated this as a Moderate. So, this wasn't held back by staff, necessarily. And we use an algorithm inside of our Measure Evaluation Criteria to determine the level of confidence that we have in the Evidence. And that's based on what we call QQC, which is the Quality, Quantity, and Consistency of Evidence.

And the nice thing is that our little flow chart has some clear delineations that we can use as a standard and, if so, very comfortably inside of the Moderate box.

Any additional comments on Evidence? Great. So, I think we can move to the vote.

Vicky Shanmugam: Can you give us some quick instructions on how this is going to work?

Sheila Crawford: You should have the link in your email to the Poll Everywhere. You'll need - okay, we're pulling up the…

Woman: I'm sorry, where's the link? You sent an email and it has…

Adam Thompson: Close towards the bottom of it. If folks are looking for the link if you open the email, it's actually towards the bottom of the email. It's the one that actually looks like a Weblink versus just a name.

(Asaba Denwulquafor): So, we have the Poll Everywhere open for Measure 2522 and we're going to be voting on Evidence. The maximum that this can be rated is Moderate, Low and the Lowest will be Insufficient.

Woman: The maximum is Moderate?

(Asaba Denwulquafor): Moderate, Low, or Insufficient is how you're going to rate it. You forward the Poll Everywhere link.

Woman: Yes, I see it. But, how come Moderate is the top choice?

Woman: It's based on…

Adam Thompson: You've got too many (unintelligible).

(Carol): So, this is (Carol) from NQF. It's based on the Evidence Algorithm that we follow. So, based on what is submitted, the highest possible is Moderate.

Woman: So, I guess if you guys are already predetermining it, why are we voting?

(Carol): It's not predetermined. You have the option of Moderate, Low, or Insufficient. You don't…

Woman: We don't have the option of High?

(Carol): Not for this Measure, no.

Adam Thompson: And the reason is because the - our Algorithm - which we can bring up if you'd like to see it. It stipulates what the rating can be based on. The number of studies and the strength of those studies. And the direction of those studies. So, once those have been evaluated by staff and you can look at it yourself, if you wish.

Woman: That's fine. It would be malpractice for someone okay. And there's good evidence that these drugs for patients with substantial risk of tuberculosis. I guess I'll just agree to disagree with the results of your assessment.

Adam Thompson: Okay. Well the nice thing is, that regardless of whether it's High or Moderate, it's passed. And it's just like whether or not you graduated from a University that many people haven't heard of or some fancy school, you're still a doctor, right? So, I agree that there may be some strengths of studies that are

overwhelmingly powerful that may not align perfectly with our Evidence Algorithm, but yes, thank you.

(Carol): And the other thing I would add is, we are not evaluating the degree to which the evidence that's out there may meet the measure. It's what's presented in the submission. And so, that's what we base, the staff based their review on. So, we understand that for this particular issue, there may be a strong Evidence base, but this is based on the submission that was submitted to NQF.

Man: So, the system has always (unintelligible) that and done this for a long time. So, the question is, can the committee question the staff evaluations?

(Carol): You can, but what we think would be helpful is to walk you through the Algorithm. So, the Algorithm is based on the criteria. So, these are not arbitrary, you know, criteria that we brought up. This is NQF's Measure Evaluation Criteria that you're using to evaluate the measures.

And this is the degree to which the Measure meets the criteria. So, perhaps you start at the beginning (Hual) and (Asaba).

Adam Thompson: Would it be possible to put this on the other screen so that (he) can see it? Yes. All right. I'm going to go ahead and walk through the Evidence Algorithm with you just so you can understand how NQF evaluates Evidence.

So, the first box is really determining whether or not this is an Outcome or a Process Structure, or Intermediate Outcome Measure. Since the answer to the question is, is this an outcome measure or patient reported outcome measure? The answer to that is No. So, that places us down to Box 3. And if there's a different type of evidence that we expect for Patient Reported Outcomes and Outcome Measures, so this leads us to our QQC Section.

So, this Box number 3 is where we are now. And for Measures that Access Performance on Intermediate Clinical Outcome Process or Structure, it is based on a systemic review and grading of the body of empirical evidence when the specific focus of the evidence matches what is being measured.

So, what we're looking for is the answer to the questions that you'll see below there. Is the evidence about something other than what is measured? Empirical evidence submitted but not systematically reviewed? Based on expert opinion? No evidence because it wouldn't be - won't be studied? Or distal process step is not the specific focus of evidence?

So, the way the staff maps this one - (Gail) do you have the actual staff form? Yes, well, what box did you go to next on this one?

(Gail): Sure, so it goes from Box 3 to Box 4 and then it goes to Box 6. So, if you see from Box 6, the highest you can rate it is Moderate to Low.

Adam Thompson: Low is not Pass. Moderate is Pass.

Man: Just out of curiosity, I saw blood pressure screening was listed as an example in the Algorithm. What would the level of evidence be for blood pressure screening?

Man0: Oh, that's a good question.

Man: That you ever would High level for this, I guess.

Adam Thompson: Yes, yes, we do.

Woman: Well wearing a parachute when you jump out of a plane.

Sam Stolpe: Well that's a great example, but I'm a bit surprised people are questioning that we can have something that's malpractice if we don't do it. But the quality of evidence is listed as moderate. It re - I just went through multiple guidelines, I mean, for all sorts of topics. There're tons of moderate evidence that we make recommendations based on because there aren't randomized trials of jumping out of airplanes without a parachute. And yet none of us would do it.

So, there's, I think it's just not uncommon. When I looked through this particular measure, I remember reading through it. Most of the studies submitted actually weren't in rheumatoid arthritis patients, they were other patients receiving immunosuppressive therapy, which is very relevant to this particular Measure, but just walking through the Algorithm doesn't take you to that High level of Evidence.

So, just - I think the important thing here is Moderate passes and we move onto the next category.

Adam Thompson: Any other additional comments or questions? (Unintelligible) are vote (unintelligible)? Nineteen is the total we're looking for.

(Gail): Okay. So, voting is closed. (Asaba) will read out the results.

(Asaba Denwulquafor): So, for Measure 2522 Importance to Measure Report for Evidence we have 18 people voted for Moderate. So, that's a Pass and we'll continue.

Sam Stolpe: Next up, we have Performance Gap. I would say, does anyone else have anything they would like to add related to Performance Gap or do they feel comfortable moving to a vote?

Man: Actually, I have a question about that. So, as I recall when I looked through the Measure, I don't have it now because we didn't talk about it. Performance Gap the last report was in the 50%-point range and then it went up to around 85%. I'm curious, since, you know, there seems to be strong consensus that tuberculosis screening should be done for any patient on these disease modifying agents, what's the signal to noise ration?

I mean that 15% that are failing, do we know they're really failing? Or is it just an inability to capture the data out of the electronic medical record which I think, you know, as you mentioned,, documentation of PPD skin testing, and other things that are often done in employee health and lots of different places, that sometimes might not flow easily to electronic medical records. So, are you constant that they aren't doing it 15% of the time?

Lisa Suter: Yes, because we've actually done rigorous validation of the elements for all of these measures with chart review. So, where electronic data is polled and a clinical abstractor goes into the medical record and verifies that the data is both reliable and accurate. And I think your question about signal to noise is going to be addressed under Reliability. And the answer is yes, but I'll address it under Reliability. I'm getting looks from Sam.

Sam Stolpe: Any other additional questions or comments related to Performance Gap? All right, we can move to the vote.

(Gail): Okay, just a reminder. Everyone should be logged into the Poll Everywhere. We have 20 people now present on the committee, so please make sure you're logged into the Poll Everywhere link.

So, we should be voting on Gap. So, we'll advance the Slide to Gap.

(Asaba Denwulquafor): The options for Gap are High, Moderate, Low, and Insufficient.

(Gail): I think we're missing one more vote. Okay.

(Paul Metchers): (Paul Metchers) speaking. Can you again, go over the Pass, Differentiate the High versus the Moderate again? And how many people need to vote High versus Moderate?

Woman: High and Moderate are added together to make a Pass. And Low and Insufficient are added together to make a fail.

(Paul Metchers): That's what I thought. So, whether the staff rate is High or Moderate for voting purposes doesn't make a difference.

Woman: Right. You should vote whether you think it's High or Moderate. Doesn't necessarily…

(Paul Metchers): Great.

Woman: …but then we add those together.

(Gail): All right. And do you need another second, or are you having some technical issue? Okay. Well we'll go ahead and close it and you should be able to catch up with the next line. All right. I will pass it to (Asaba). Okay, great. I'll pass it to (Asaba) to read out the vote.

(Asaba Denwulquafor): So, for the Gap in Performance, we have 7 Committee Members voted for High, 12 voted for Moderate, 0 Insufficient and 0 Low. We'll go ahead and lock the poll now.

Sam Stolpe: Thank you. So, now, Scientific Accessibility.

Katherine Gray: The work - and I'm not quite sure what the group would like to do with this. There was a lot of comments about the Reliability, you know, being Insufficient, which it was when we reviewed it. So, I'm not quite sure how the group would like to handle discussing that. The comments and so on, were lots of issues and concerns, you know, including, and I still, you know the (unintelligible) of the group, you know, what that might mean in terms of Reliability, those kinds of things.

One question I was going to ask. Is it Lisa, is that right? Yes, I was going to ask you, you gave the average number. But do you have a median number in there? Well okay, you don't have to answer that immediately, but my big question had to do with throughout the measure and submission, it talks about the fact that there was, you know, increasing things but then there was decreasing, you know, numbers and trying to explain that has to do with demographics shows some - I'm just trying to gather up - I mean the thing about Reliability is, that you want to know that whatever you put in place a year or two from now, it will be measuring the same stuff and that we don't have to be thinking of why it's different. You know, at least (unintelligible) we know what we're doing.

So, I was just curious as to, you know, do you believe that this is, you know, getting at the things across, whatever all the changes, you know, more migrants? More people from Africa? You know, those kinds of things that may be making a difference in terms of what, you know, our and/or the different socioeconomic groups, if they are getting the same treatment? Or, you know, what is it that is going on here? And is it, you know, something you can do reliably?

Lisa Suter: So, thank you - my apologies, thank you to the NQF staff who pulled up the document that we shared. This is Individual Level Clinician. The distribution of the Reliability Scores using the same signal to noise approach that we used for the practice level. There are over 500 clinicians in this. The median which you asked for is .84. This is all volumes of providers.

The table below shows that if you truncate only providers that have a minimum of 10 patients, and you'll see that that bottom (quartile) is less of a floor - I'm sorry, a threshold and you see good Reliability all the way down to the minimum Reliability.

The other thing that came up, and I guess Dale you were asking a little bit about this and I apologize for not directly addressing it. But when you pull out lower volume providers, you still see the same range of performance, so I don't think it's a volume issue in terms of performance. Trying to get at some of your questions, Katherine, the - we're not capturing whether or not the tests were positive. We're only capturing Best Practices, were the patient's screened? And we have a 12-month lookback.

I believe it was erroneously noted as 6-months. There is a 12-month lookback to make sure that they haven't already received - that they've received the testing or that they've been treated with a drug appropriately if the test was positive. So, we are capturing those patients, but we're not necessarily capturing - we haven't looked at the definition of whether or not you're successful in this measure because you're treating someone who had a positive result. We don't really know the underlying prevalence of changing demographics.

There are disparities in rheumatoid arthritis care and so all these measures serve only to increase awareness of cost different patient population. I'm not sure if I'm entirely addressing your concern?

Sam Stolpe: One second, real quick. Just a process point. Because I know some folks are unable to see the screen up there. So, they just emailed out to everyone in the committee the link to the Webinar. And if you actually sign into the Web Platform, then what you're seeing on these screens will be mirrored on your computer screen.

Woman: And we've also emailed out this PDF file so that you can look at it directly on the computer, as well.

Sam Stolpe: Okay. Katherine can you…

Katherine Gray: This is Katherine. Is what we're looking at here - isn't this what was in the original submission? Or is it changed to the be the (team)'s (unintelligible)?

Lisa Suter: So, what was in the original submission was at the practice level. And what you're seeing here is at the individual provider level. So, you are now seeing both Reliability at Individual Clinician and at the Practice Level. And I, first of all, thank you to the Committee for considering this additional data and thank you to the NQF staff for bending over backwards to make it available.

And we will proceed going forward always doing it at the Individual and Group Level. But this fills that gap that was flagged during the Review Process of the missing data.

Katherine Gray: Yes, and one more follow up question. For those people who are in - because they're with - and I don't have it here in front of me, but there was one study

that talked about, you know, the lack of treatment if you were of lower socioeconomic status. So, that would suggest that either, A, they didn't get the TB screen or nobody treated them, or, you know, there's something wrong here in this picture. Can you kind of elucidate us on that?

Lisa Suter: So, since we think the data, we're pretty confident that the data is accurate, I think there are just really variable practices in terms of how people manage and initiate disease modifying therapy just by the Evidence and Best Practice Guidelines that exist.

Katherine Gray: That being like that the people are poor and they can't afford the copays, or what do you think is the thing going on. I mean, I realize we're speculating, but…

Lisa Suter: So, and I have to defer to (Tracy) on the line, whether the Measure captures the Clinician ordering the test and the patient not getting it because of whatever reason, fear, finance, you know, any number of drivers of noncompliance. We do capture latent - histories of latent TB, but I'm not sure whether or not we capture the fundamental issue of whether or not you ordered the test and somebody didn't get it. So…

Katherine Gray: I wasn't so much focused on the test. I mean, there's that piece, is the missing piece, but it was more focused on that they don't get the treatment. So, that was sort of, and so it was like leaping over the test. Like did they not get the test or did they just not get the treatment.

(Tracy Johansen): So, together, we have - I mean that's why we have a suite of measures. So, this measure really is capturing, are you appropriately screening your patient's.? And then we have a disease modifying demar - antirheumatic drug measure that's checking, are your patient's on disease modifying agents.

I know there's a lot of concern about the disparities in this group and the ACR is also concerned about the disparities.

We're unpacking our data. They're trying to understand what we can look at. We've done some preliminary analysis, with if you've ever been on Medicaid as a crude metric of socioeconomic status. We know that out comes our worse for those populations.

But we're - and we know from literature that that reflects access, it reflects provider behavior. It reflects a number of different things that contribute to that disparity of outcome and one of which may be that next they may not be getting the drugs.

But this measure is focused on the screening. We're working as an organization to try and focus on those disparities and to actually make them transparent as part of the Dashboard. And that's one of the things we're trying to work on right now.

Katherine Gray: Any others on the committee?

Vicky Shanmugam: So, my understanding of this Measure is that what your denominator is, anyone receive biologic (demand) of that population, did they receive testing for TB? And I have one concern in regard to the list of biologic demands in there. So, I would know it and I think this might be where your Gap is coming from.

There isn't actually a recommendation in the FDA Guide to screen people who are getting Rituximab. And I think, Rituximab should be removed from that list because it isn't actually a concern with Rituximab, it's a concern with

everything else. And so, you know, I think the majority of providers and particularly those in private practice whose patients are getting TNF Inhibitors through an insurance program of any kind, cannot proceed through that process of prior authorization without having proven, literally on paper, that somebody's had TB screening.

So, my guess is, your Gap is your Rituximab group and I think that group should be removed from your population.

Lisa Suter: This is a great flag for us. And yes, we had a conversation just last week. It's a great catch and we will remove Rituximab patients. We're actually working right now to expand this measure so when it comes back to you in the future, it will not be limited to rheumatoid arthritis patients. In which case the (unintelligible) will also be on that list as an additional (Unintelligible) agent that's not appropriate for TB screening. So, thank you for that recommendation.

Sam Stolpe: (Don) and then (Lindsey).

Don Goldmann: I'm going to say this is probably going to seem like a stupid question, but I may have missed it. What type of testing is specified, if any? Is this Interferon Release Assay Interferon Gold or is it a skin test?

Lisa Suter: Both. And it - either will qualify.

Don Goldmann: And so, if it's a skin test, does that require documentation that it was read properly or read at all?

Lisa Suter: It requires documentation that was read. Can't comment on whether or not the reading of the test was appropriately done.

Don Goldmann: This is actually a huge issue, as you know, especially if you're working in the Navajo Nation and somebody's gotten a skin test and then they go back after having travelled 50 miles, but in a rented vehicle of a friend of a friend. And they they're going to come back and get it read and I'm just wondering how reliable is the reading part of that measure and what about that whole issue of it being read properly.

I've had my skin tests read incorrectly several times by competent nurses, so, what do you do about that?

Lisa Suter: I mean what I will just say is, these are the clinicians that are responsible for prescribing those medications and caring for those patients and so, they have the responsibility to appropriately evaluate the testing that's appropriate for the medication they're prescribing.

If they don't feel confident in evaluating those tests, they only need to document that someone has appropriately screened the patient, right? We just need the results. We don't need to know that it was the clinician prescribing the (unintelligible) that tested for TB. It could have been done by the Primary Care Doctor. But as long as we have a documented result that allows us to indicate that the test we performed or that there was an appropriate reason for not screening them, then that's what we need.

Dan Goldmann: And not to be get too deep into the weeds but does the guidance allow self-reading? Or is that to be read by a clinician?

Lisa Suter: Again, we don't go down to that granularity.

Dan Goldmann: Yes, I think then that raises significant potential for false readings, either overtreatment or under treatment. So, in publicizing the measure, I think it would be important to give some guidance around this, not just to leave it to the clinician, especially considering underserved populations and their assets of appropriate reading.

If I were a doctor in Navajo Nation who ordered this and I'd say ask the patient to come back in 48 to 72 hours for a week and 70 miles away, is not realistic. So, I would say read it yourself and let me know.

Lisa Suter: I will note that in general, because of the challenge of brining someone back for a skin test read, the rate of skin testing in our data sample is declining and I anticipate that it's really dying out. So, the FDA is now working on guidance for its tuberculin testing, so we're hoping for some better regulatory oversight of the tuberculin testing in terms of blood testing.

But I don't anticipate that this is going to be as much of an ongoing issue just because utilization and cost and convenience really drive away from PPD testing.

Sam Stolpe: Yes. Lindsey:

(Lindsey Bachford): So, taking us back to where I think we need to go, which is voting on Reliability, it seems that since the Developer has provided this addition, it goes from Insufficient to a discussion over whether the evidence becomes Moderate or Low. I'm just looking right here. I'm assuming really our discussion should be whether we think it's Moderate or Low given the data they've now provided?

I mean, I think, I'm kind of in between on it with this data. I think, you know, the Individual Level data is consistent with the Practice Level data. I do think some of the specifications as discussed earlier, make it a question of whether it's Moderate or Low.

So, yes, so I just wanted to see if we can focus the discussion on other people's thoughts between Moderate and Low to get us near our vote.

Sam Stolpe: And a hush falls over the room.

(Lindsey Bachford): Am I correct that Low is…

((Crosstalk))

(Lindsey Bachford): Is low meaning a Not Pass?

Sam Stolpe: Correct.

(Lindsey Bachford): So, that's the discussion.

Woman: Lisa, could you give us a little more detail on the two analysis that are here? In particular if the sample sizes are different and the minimum Reliability is quite different (unintelligible) the two. One is .06 and one is .72 and I would have expected, is the bottom-line individual level and the top one is not? What are we looking at here?

Lisa Suter: So, they're both Individual Level. One is not volume restricted and the other one is restricted to providers that have at least 10 patients. So, understandably, if you're reporting on one patient or two patients, does Reliability of that test, is really poor.

So, if you draw a line at 10 patients and say you have a minimum requirement of volume of patients in order to report this measure, the Minimum Reliability is .72, which matches NQF .7 cutoff. So, if we are happy to hear guidance from the committee on approving this with a Minimum Volume Requirement, we think that that's appropriate for scientific acceptability and it doesn't mean that we won't give practitioners information about their performance. It just means that they will not be able to report this as for example, part of MIPS from this version that's rolled into MACRO Program.

But it does mean that what gets publicly reported will be statistically reliable and more meaningful therefore.

Sam Stolpe: Just a brief point of clarification. So, this is core level reliability testing which is Developer provides this and it opens up the option for it to be rated as High, as well. And we're staff evaluating this. We will likely land on a Moderate to High rating based on the results, that the Developer has provided.

Any other questions or comments?

Woman: I was going to say, can we - I realize it's a little different then maybe just the straight up voting, but can we make that qualifier about it being - because it was part of the question too - about the minimum number? Can we say that that's how we vote, you know, is related to also assuming a minimum number of 10?

Adam Thompson: Not if it's not part of the Measure Specifications, which 10 is not an exclusion (criteria) on this measure. So, the Measure Developer would have to specify that. So, this is just supplemental data that the Developer has provided which

doesn't align necessarily with the specifications of the measure. But does give you some idea of the behavior of the reliability.

This could be really helpful for say, an Implementer who is looking to put this inside of a MIPS for example for determining what thresholds would introduce stronger reliabilities, if that's something they were looking for. But for your voting purposes, the one that you'll want to focus on is the one that is how the measure is specified.

Do you have a comment?

Woman0: Can I - I'll just add that MIPS reporting threshold is 20 cases. So, just to reassure you that the public way in which this measure is being used has a minimum reporting value of 20. And that, you know, the ACR, I'm getting texts from the staff throughout this conversation. You know we can guarantee that we will update the specifications and ensure that for reporting or accountability purposes, this measure - we would make sure that this measure has a volume cutoff of 10.

That does not, again, we want to make sure that even if you don't have 10 patients, you're still getting information about your performance. So, that's not the Quality Improvement Purposes but for Accountability Purposes, we would ensure that that volume cutoff is minimal, (so).

Sam Stolpe: Thank you.

Man: I just want to echo the point you brought up. So, basically for our purposes it wasn't required to have a minimum amount. But for all intents and purposes, (do not) report on one patient for this measure. So, when we look at reliability, we're going to assume that it's going to be more than one patient,

because the reliability obviously goes down the drain when you're just reporting on one patient?

Woman: Just sort of a process questions. So, the Measure Developer responded to the comments and went to the trouble to do another analysis which they have presented here. And it sounds like what the Measure Developer you know, would do, if he had to change the effects of the Measure to require a minimum threshold of 10, such that the measure would be reliable.

Yet we're being asked to vote on the measure in its original form and I'm just wondering shouldn't there be some flexibility in the way we vote? Because if we have to vote it on its original form, then that means the Measure Developer has to go through this whole process again to resubmit the measure and we all have to review it again, and we have to vote it again, when we could just take care of it right here and now?

Adam Thompson: Yes, I mean, I've definitely seen a vote in favor of something with promises from the Measure Developer to sort of address issues that the committee had. So, I think it's possible for us to consider the agreement here to talk about the low volume threshold when we make our vote.

Am I out of…?

Woman: Yes, I think I would like to get clarity from the Developer in terms of timing. How soon could you make the changes to the specification?

Lisa Suter: (Tracy) can you answer that question for…?

(Tracy Johansen) Yes, this is (Tracy Johansen). One thing I do want to know, in terms of the minimum volume here, I guess more of a question I would like to ask. You

know, we want to make sure our specifications are set up so that if (somebody) just wants some people to track this information on their own. Or if they only have a couple patients who might meet the denominator for this are still able to see what's happening with those patients. But as Lisa pointed out, in terms of accountability, we would not allow someone with just a couple patients to be held accountable for that measure given below their liability when you're just looking at a couple of patients.

So, my recommendation or my thought that I would take back to our group that would review this would be to essentially update the guidance in our measure so that if this is being used for payment purposes outside of MIPS, there's already the requirement within MIPS, but it has a minimum of 20 cases, that we would at least have a minimal of 10 cases before any accountability - I'm sorry, use in any Accountability Program.

Is that acceptable as an adjustment to the Measure and the Specifications?

Man4: If this is at all helpful, the staff evaluation for the top portion of this would actually be fairly high. Because the first quartile reliability at 0.73 they're saying that below that is where we start to exhibit some level of under liability. And that means that, you know, over 75% of the clinicians that were evaluated in this, we have a strong confidence that we've ranked them appropriately.

For our purposes, we would say that this is actually really solid, even at the top level. And if the Measure Developer is making these sorts of clarification points, then I think the staff would feel comfortable with that. Dale?

Dale Bratzler: So, again, I think we're getting into conversations that we don't need to. How to measure - so if I'm a clinician submitting to a registry and I have nine cases,

I want to know how I did on the nine cases. I don't want someone to pay me differentially based on that, but it's not, I don't think it's our committee's job to decide how performance measures get used.

We're supposed to vote on whether the methodology the Performance Measure whether it's got Evidence Base, Practice Gap, whether it Reliable and Valid. Those are the things that we're supposed to evaluate the Measure on.

It's not our decision about how a Performance Measure gets used. Some are used to Quality Improvement Purpose, where I would argue that the denominator may not matter and others are used for accountability. MIPS already, you know, the one that's been discussed here, already has a 20-case minimum threshold that CMS is already set up.

So, I think we need to be cautious. Your comment about Rituximab though is very significant, I think, substantial change to specifications for Performance Measure that does need to be addressed. Because that actually effects the Scientific Acceptability of the Measure, I believe so.

Sam Stolpe: Thank you.

Woman: And I think we can make that immediately, that change.

Vicky Shanmugam: Yes, so I was just keen to kind of move us along to a vote on the Reliability with the minor modifications proposed that we remove Rituximab and put the cutoff at 10 patients as the minimum reported. And then I think we should vote.

Sam Stolpe: All right. So, let's go ahead and vote.

(Asaba Denwulquafor): So, we're voting on the Scientific Acceptability Reliability for Measure 2522. The Highest level that we can vote on is High, Moderate, Low, or Insufficient. So, this is Measure 2522, Tuberculosis screening for Rheumatoid Arthritis.

(Gail): Microphone please.

Woman: And also, before I think you need to introduce yourself and let us know if you have any conflicts and that's before you vote and before you make a comment. Thank you.

ANA McAllister: Sure, ANA McAllister. And I have no conflicts. Just one clarification, Dr. Shanmugam proposed an exception. I'm just wondering if the vote that we're taking right now includes that exception.

Sam Stolpe: Yes, so the Measure Developer agreed. I believe she said we could do that right now. So, that our vote would be inclusive of that, as well as the guidance around the low volume threshold and Accountability.

ANA McAllister: Okay, thank you.

(Gail): I think we're looking for one more vote. Is anyone having a hard time submitting their vote? We're at 20. Okay. All right. We'll have (Asaba) read the results.

(Asaba Denwulquafor): So, for the Reliability portion for 2522, 3 committee members vote High, 15 voted Moderate and 2 voted Low.

Sam Stolpe: Do we have any specific comments related to Validity?

Woman: It was recommended at Moderate and there are no exclusions.

Sam Stolpe: All right. I think we can move to a vote.

(Gail): We're pulling that up right now.

(Asaba Denwulquafor): For the Validity for the Highest is High, Moderate, Low and then Insufficient.

(Gail): Okay, I think we're waiting on two more votes. Please submit your vote if you haven't already. Okay. We're still at 18. Two more people. Oh, one more person. I don't know, Jeff, if you have submitted your vote for validity? Okay. All right, well we're going to go ahead and close that. No, we're still above quorum. Okay, (Asaba) can you read the results?

(Asaba Denwulquafor): So, under the criteria for scientific acceptability and validity, we have one committee member voted high, and 18 voted moderate, for 2522. So the measure passes.

Sheila Crawford: Thank you. Catherine, feasibility?

Catherine MacLean: We'll keep this simple. The feasibility was rated at moderate. And other than one comment by one of the panelists, that they asked the question about, does using the (RISE) registry suggest that, you know, that it does seem feasible? So, you know, there are some drawbacks, if only 30% of the physicians who were ordering the drugs, the biologics, that, you know, is that really feasible? But no one came up with anything more negative than that. That's it.

Sheila Crawford: Thank you. Any additional comments? Oh.

Woman: I was just going to comment that I feel like the issue really is, in feasibility, is this structured data field, as best I can tell, right? Is it easily pullable from whatever (regency) you're using, whether it be Epic, (Sona), one of the others, right? And my impression is it would be, with the caveat that sometimes the PPD testing is not quite documented the way people would ideally want, with the move I think generally in practice towards the interferon release assay type testing. My impression is that this wouldn't be a problem in terms of feasibility.

Sheila Crawford: Thank you. (Don)?

(Don): Yes. So I'm not objecting to the feasibility ranking and statements, but I'm going to continue to make statements about the holistic nature of healthcare. And pulling data from employers, hospitals, schools, primary care practices is arduous, and there is conceivably unintended consequence of (unintelligible) to be able to repeat testing because that's just too difficult to do. So, would a person coming in and soft-testing, that I was tested in my hospital because I'm an employee of the hospital, would that count? Or would they have to produce documentation? Would the specialists have to do this? Why are we only dealing with registry pulled data? It's just, I mean optically, it's a typical top-down measure that doesn't really relate to how people practice in the real world. That said, I'd probably vote okay on it, but.

Woman: I'm going to comment on that as a clinician in practice. So it's pretty normal in most of our EMRs that if you are wanting to document, say, a vaccination done outside, a PPD done outside, you have a mechanism by which you can do that in a (unintelligible) right? So the question is, are the providers doing it rigorously or reliably on every patient? Probably not right now. But if they

were told they were being measured on it, they probably would do a better job.

And I think for those of us who are prescribing these medications, I can tell you, we have to provide documentation that the patient's had testing, or the patient won't get the drugs and the insurance. So I'm actually not concerned about that. We got to be safer. If they've had it done at their school, we get that piece of paper and scan it in the record, yes, then it's a scanned document. So you would have to then have some mechanism for that doc to click the necessary box to make that now a structured data field. But I'm not worried about that being feasible. I think it's very feasible to build that into the system.

Sheila Crawford: Thank you. (Wendy)?

(Wendy): So my comments overlap I think both feasibility and usability. And I think because of the feasibility of the documentation of the PPD, I think it does create a little bit of an unintended consequence of overuse of the immune-assays because it's easier if you have ordered and gotten electronic results back in an EMR to meet your measure than to do these two set process of a PPD that's often done on a paper or not in a structured data field.

So I don't know if that way is the overall usefulness of the - of the measure, but I think the challenges in feasibility could potentially assist a little bit of what it's done in practice because of ease of meeting the measure. Not being a rheumatologist, I don't know that I can weigh in exactly, but I know we see this in other measures were the documentation burden of meeting a measure is too much, so it's easier just to repeat something than to go through the process of documenting an EMR.

Man: I think you make it seem like a piece of cake and I'm sure in your practice it is, but in general in American healthcare, it's not. And I wonder, what did the patient representative say in the construction of this measure and thinking about its feasibility and use?

Woman: I guess I was going to make comment or respond to your comment of use. I can tell you that that is probably is true. I can tell you that in real life people do use now the interferon gamma release assays rather than other methods because of ease of use and ease of documentation and ability to prove to the insurance companies. I would say there is a degree of both positivity rate and we are recognizing that. This measure is not testing that. This is not what we're being asked to comment on.

What we're being asked to comment on is, is it okay to prescribe these drugs without having tested? And the answer to that question is no. The question now we're being asked about on in this specific place is, is it feasible to measure how well the performance of the physician or the provider is on this testing? And the answer is yes, it's totally feasible. We should be documenting that in our medical records. Right? This is a test that is required in order to prescribe these drugs safely. We've already had that comment made. And yes, we should be able to measure it.

So I would - I'm less concerned than others on this.

Man: (Do we have) what the patient, who's the object of this testing, had to say about the measure and the possible burden on the patient?

((Crosstalk))

Man: Let me ask it differently. Was there a patient representative in your measure construction (committee)?

Man: I mean, I can just comment only on the fact that we had a similar thing happen in HIV, right? I mean, I'm a hospital employee and I'm a person who's screened regularly for TB. And we moved almost entirely to QuantiFERON, not because of a documentation issue. It was getting the PPDs read, right? It was getting that result.

And when I tried to submit my HIV screening for TB through the hospital, it's not that they can't take that. They made an internal choice not to take that, right? So I don't think it necessarily speaks to the measure so much as it does like hospital practice, right, which is kind of like I think a different issue. I would agree with you, which is not so much - like that's not kind of what we're looking, but I - does it create more people like getting a QuantiFERON? I mean, I can tell you in HIV it absolutely did. But at the end of the day, we also were able to show that we had way better results on understanding the rate of TB in our population because we actually had results.

Sheila Crawford: You made a point, let's move on. But I think (GenQF) should seriously consider requiring patient representation on all measures submitted to NQF.

Woman: I mean, I am a patient representative, I do not have rheumatoid arthritis, and I mean, this is not a disease area that I know that well, I have diabetes. So I mean, any kind of test does add an additional burden to a patient, I mean, whether it's important or not is a different issue, but there's additional burden. However, if you're taking a disease-modifying agent such as some of these auto-immune drugs, you want to make sure that the measure is - that, you know, you're taking the drug for the right conditions.

So I mean, you don't want to take something that's inappropriate, and if the test would tell you whether or not appropriate, then that would seem to be a reasonable burden.

Man: It's not up to us to decide what's a reasonable burden for patients. It's the last time I'm going to bring it up today. But we do not as physicians, providers, regulators, measurement developers, decide what's a reasonable burden for patients. Patients participate in the process and always should in this day and age, just like moving to a better blood test as opposed to a stupid, antiquated skin test that nobody knows how to read is a good thing to do. It's a good thing to have a patient on a measurement development (group).

Sheila Crawford: Yes, go ahead.

Woman: So I apologize. I thought you were asking the patient representative on the committee, not the measure developer. So, yes, we had had patients through development, not only do we engage patients in the guidelines that inform these measures. They have - they are involved all the way through the process of guideline development and measure development.

And I just want to reiterate that the measure captures both. So, whatever testing process the provider and the patient and patient shared decision-making decide is the appropriate test is in this measure. So, and my experience is yes, quality measures do drive a lot of practice, but so do other influences like cost and getting patients back for, I will just say in the VA system, they switch entirely to interferon-based testing because the - just the logistics of bringing that number of patients back for skin testing was just, and employees, was not feasible. So they just moved, you know, and that had nothing to do with quality measurement, it had everything to do with just, you

know, feasibility really. So - but just capturing that the measure captures both and patients were involved in the measure development.

Sheila Crawford: Any other questions or comments about feasibility?

Man: Thank you for that, and that's very reassuring and I'll shut up now.

Sheila Crawford: All right, let's go ahead and move to our vote.

(Asaba Denwulquafor): All right. So (unintelligible) active for feasibility criteria, we're still within measure 2522. The options are high, moderate, low, and insufficient.

I'll go ahead and lock the poll. We have 20 members who have voted so far. Eight voted high, 11 voted moderate, and one committee member voted low.

Sheila Crawford: Thank you. So we'll move to usability. Oh, Catherine, your microphone.

Catherine MacLean: Yes. It looks as though the usability passed and was voted as a first use to assess, and then usability was voted moderate. And generally, no comments that I would consider negative. There's the continuing comment about not everybody participates (unintelligible). That's it.

Sheila Crawford: Any additional questions or comments related to use or usability? All right, I think we can move to the vote. Two votes.

(Asaba Denwulquafor): The polling is now active for the use criteria. Committee members can either choose pass or no pass.

Woman: Okay. I think we're just waiting on one more vote.

(Asaba Denwulquafor): So we got (22) votes, and all committee members - and all 20 members voted pass on the use criteria. I'll go ahead and activate the usability criteria. Polling is now active for usability, still within the measure 2522. The options are high, moderate, low, and insufficient. So we have 20 votes in. Six voted for high on usability criteria, and 14 voted for moderate. So the measure passes its criteria.

Sheila Crawford: Now we move to overall suitability - yes.

Woman: And the last one, the related or competing measure does not apply. I don't know what that means. We vote (we don't).

Sheila Crawford: No vote.

(Asaba Denwulquafor): The last criteria for 2522 is now open, overall suitability for endorsement. For this measure, the options are yes or no committee members. So the final voting criteria for 2522 overall suitability for endorsement, we have 20 votes. Nineteen committee members voted yes, and one person voted no. So the measure passes.

Sheila Crawford: Great. Thank you so much.

So I just want to let - give folks just a time check here. We're about 20 minutes behind where we kind of need to be to keep our schedule for today. It's not a problem, we encourage discussion. I just wanted to kind of put that out there.

So now what we'd like to do is move on to our next measure, 2523, assessment of disease activity, rheumatoid arthritis. Might I ask our lead

discussant, (John), or I guess measure developer if you would - nope. We're good. Right. If our lead discussant can go ahead, briefly summarize the measure, and then move into specific information about evidence.

(John): Great. This should go much more smoothly, given the fact that most of the concerns were already addressed. So, thank you, Catherine. You made my job much easier.

This is 2523, rheumatoid arthritis assessment of disease activity. The description is the percent of patients greater than or equal to 18 years old with a diagnosis of rheumatoid arthritis, and 50% or more of the total outpatient encounters in a measurement year have the disease activity assessed utilizing one of the six standardized measure. It's a new measure. It's a process measure.

And the initial discussion on evidence, the NQF staff reviewed the evidence and gave it a moderate rating. Based on the algorithm, it did include the American College of Rheumatology 2015 guidelines, which did include systematic reviews. They did look at the quality consistency and quantity of the data, and actually used the grade methodology to rate the evidence. So they met the criteria for a moderate rating for the evidence.

Sheila Crawford: Thank you. (Bill)?

(Bill): I have a question about the outpatient visits. So if a general internist is seeing their patient with rheumatoid arthritis or other chronic disease management, and they discussed the rheumatoid arthritis during the visit, it's put in as an ICD 10 code. That internist has five visits. The patient sees the rheumatologist for two visits. The rheumatologist does the screening; the

general internist doesn't. So that would be - so my question is, will those visits by the general internist be counted in the number of visits?

Woman: So, (Tracy) can correct me. Right now it's less of an issue because right now it's a registry-based measure. So we don't have primary care clinicians that aren't also registry participants measuring this. Right now it is - the definition of the denominator is based on the rheumatoid arthritis coded visit. Certainly we can explore opportunities for acknowledging whether or not it was a primary active issue in terms of doing that, but that's one of the reasons that we picked a lower bar for capturing disease activity, of 50% of the visits, because we recognize that there is some - there's a mix of appropriateness for different (encounters) for rheumatoid arthritis in terms of when you should be capturing disease activities versus not.

I think in the current iteration of the measure, in its current implementation and its expected implementation by (MIPS), which is, while it's a mandatory program, you voluntarily select the measures that you submit, we don't see that this is an issue. But (your fragging it) is really important for us going forward as we continue to think about evolving the measure and expanding it beyond the registry. Thank you.

(John): I think I didn't mention the committee comments. And in general they were supportive of the evidence. There was one comment about it may be difficult to implement, but I'll leave that for feasibility.

Man: I'm sorry, that was an excellent question. So if you're an internist and you have a diagnosis of rheumatoid arthritis, but you're not really treating it, are (you deemed) on this measure if it goes forward?

Woman: Only if you code that diagnosis as something that you are actively caring for during that visit. Now the measure does - allows for flexibility in which providers are actively caring for rheumatoid arthritis and collecting disease activity. You only have to collect disease activity at 50% of the visits.

Man: Right, but that's an untoward consequence. So if you're treating him for high blood pressure and you put rheumatoid arthritis as the fourth diagnosis, you have to be darn sure you're documenting that to evaluate that, or less you're going to get (deemed), or you have to tell the - all the internists in the United States not to put it down, even though you listed it at some point for whatever reason.

Woman: So, two issues. One, in its (unintelligible) (limitation), it's a voluntary measure from (MIPS). You have to select reporting this measure. So, no internist is going to have this measure reported on their - for them without their participation. And right now it's implemented only in a registry where this is avoided, because internists, you know, if for example, for - there are multi-specialty practices that are in our registry, and the registry vendor and the practices have worked together to identify the providers for which encounters are (pulled), so we already have a situation in which this is managed successfully, and we do not have dilution - excuse me - from outpatient providers. We completely agree that if - as we're moving to a fully electronic (ECQM) where we have all standardized fields and this is nationally implemented outside of our registry, we will definitely need to address this. I really appreciate you guys making…

Woman: Yes, I had similar concerns actually because this measure and the next one that we're going to discuss in regards scalability, so I agree. I think that as you delineate it and develop it within a rheumatology practice community, it looks very different than when you try to scale it to the whole country. And I

think that's the major concern I have with both of these measures, was the scalability.

Woman: So, comments, important comments, and we'll them. I think something to keep in mind though is that not all patients with rheumatoid arthritis who are being treated, even with biologics, are treated by rheumatologists. And so, in some parts of the country it is the primary care doctor who is taking care of the patients. So, you know, this measure, you know, should certainly be applied to those doctors as well.

I do think that there's a challenge though to (unintelligible) it out if it is maybe primary diagnosis. And, you know, right now this is being, you know, supposed as a registry measure. So I think that, you know, with unintended consequences, would not occur in this situation. But if another measure were to be proposed, you know, perhaps based on kind of a broader electronic health record for example, there would, you know, be - I think there'll be a different measure and there need to be specifications to make sure that it was the treating doc that was being measured. But I think that primary care doctors need to be accountable as well, and it's very important.

Woman: I would say these points are valid, right? Because in reality, in practice, if I see someone with RA and they also have diabetes, I mark the diabetes code and I kind of give a brief summary of how it's being managed, but I don't - I'm not managing the diabetes but it's comorbidity to my other underlying disease, right, (unintelligible) all of those things. You note them. It's a good practice, right?

And so I think it becomes very difficult to delineate what's your, you know, your primary code obviously is the primary rheumatologic condition in our case, but I think that for an internist, they really are managing the patient with

these multiple diseases, right? So I do think they're going to get (dinged) if this became a quality measure that they were - that everyone seeing a patient with a diagnosis of rheumatoid arthritis, remembering that's 1% of the population, and yet they're not doing disease activity scores, I mean, they're not going to (meet) that. I really don't think so.

Woman: I mean, I think (Cathy)'s point is spot-on in terms of you can delineate primary versus supporting codes or secondary codes. So I think that's the best way to manage this. Right now it's not an issue so we're not delineating, because we feel like it's much better to capture for our rheumatologic member population. But, excellent points for us to take as we move forward with other iterations of this measure.

Sheila Crawford: Any other questions or comments related to evidence?

All right. I think we can move to the vote. Did you have another comment? Okay.

(Asaba Denwulquafor): Polling is now active for evidence measure 2523. The options are high, moderate, low, and insufficient. We're waiting for one more vote. All right. So with all 20 committee members who have voted. We have five members voted high on evidence, and 15 members voted moderate. So it passes on this criteria.

Sheila Crawford: Thank you. Any specific comments related to performance gap?

Man: From the committee comments, it was just (RISE) doesn't collect the social disparities data. They're talking about doing it in the future. They did cite a different registry which did show racial disparities relative to collecting data on this measure. But they themselves did not.

There was a significant gap. Again the data showed an improvement in scores. And then they dropped at the 2017 measure, which they attribute to the early adopter phenomenon.

Sheila Crawford: Any other additional comments on performance gap? Okay, we can move to the vote.

(Asaba Denwulquafor): Polling is now active for measure 2523, rheumatoid arthritis assessment of disease activity. The options are high, moderate, low, and insufficient on the performance gap. Waiting on one more vote. The voting results for 2523 performance gap. Five committee members voted high, 14 voted moderate, and one committee member voted low. So the criteria passes on that.

Sheila Crawford: Thank you. Moving on to scientific acceptability, beginning with reliability. (John)?

(John): We don't have to go into the discussion again of - I don't know - I'll ask the developer. Did you do the same level of individual and group practice analysis that you did for the prior measure?

Woman: We did. They're pulling it up right now. Again, thank you (both of you) guys and to the staff. You'll see better reliability but similar issues where, when you're looking at all providers, the minimum reliability is lower. And if you scroll down to providers with a minimum of 10 patients, your minimum reliability is excellent.

(John): Yes. So these are even better than the prior measure in terms of reliability. In terms of the committee comments on reliability, it was just the same question

that was brought up previously about that this isn't structured data and that you're having to search through the EHR to find it. But other than that, no negative comments about it.

Sheila Crawford: Thank you. Any additional comments on reliability?

Okay, we can move to the vote.

(Asaba Denwulquafor): So now voting on scientific acceptability, reliability, for measure 2523, rheumatoid arthritis, assessment of disease activity. The options are high, moderate, low, and insufficient. We're still waiting on one more vote. The result for this poll. Eight committee members voted high, 11 voted moderate, and one committee member voted low, on scientific acceptability, the reliability criteria, measure 2523. The measure passes for this criteria.

Sheila Crawford: Thank you. Validity?

(John): The NQF staff gave it a moderate rating, as they did (phased) validity testing and actually used the standardized format for assessing (phased) validity, and it got a score I think of 9, which was quite good.

And they also did a (campus) score for comparing the data extraction by a trained reviewer compared to pulling it out into the registry. And that score came out fairly good too. So it seems to me the standard for a moderate rating on validity. In terms of comments, no negative comments about validity, other than the ability to report.

Sheila Crawford: Thank you. Any committee comments or questions on validity?

Woman: …in the (RISE) registry, how are you measuring this? Can you just give us a quick summary of how it pulls out? Because in real life, these are complicated calculations that get done. These are disease activity scores that involve the patient has to have had a lab test usually, right? So it's a measure, number of active joints, plus incorporation of a logarithmic equation involving the (said rate) of the (TRP), usually. And so if you could just…

Woman: So it varies a little bit by practice at this point. Where possible and with all of our new and more recent practices that have joined, we pull all data elements that are used to calculate the score, as well as either the score that's calculated by the practitioner or we calculate it for them. But in some, you know, there are some practices where they just have the summary score and we don't have all of the data components, so we're evolving everyone to a situation where we're pulling actual data elements, because we actually are trying to create a disease activity outcome measure.

A surprising proportion of people collecting this (arc), we are collecting all of the data. So we have a robust dataset to start with for measure development. But there are still some practitioners that are documenting that they're collecting a standardized disease activity assessment, but we don't actually know what the instrument is and what the disease activity level is. And that's, I'd say that's about 20% of practitioners across the whole registry.

Woman: And so for real life, and I'm just thinking about, again, scope of really pushing this into a practice-based performance measure, how you're addressing people where they've been seen in the clinics, so they'd be eligible for the measure, they've been seen twice, right, that was your requirement, but they never got labbed, right? So they're missing a component of the disease activity.

Woman: So we accept a number of different disease activity measures. Some require labs, some don't. Some EHRs have integrated labs, some don't. Our (EHA) vendor works with those EMRs, either pulled from the EMR or pulled from the integrated or associated lab. You know, we've worked with actual labs, vendors in terms of pulling all this data. So there's a huge amount of effort backend when we onboard a practice, to make sure that we're getting all the data flow that we need.

I think going forward, the reality is we as a specialty need to be thinking about how to do this, you know, getting back to the patients, right? So, a lot of the ways that practitioners collect disease activity right now is driven by what works for us, the practitioner, and not necessarily what works for patients.

So, one of the things (RISE) is working on is creating a patient portal and ways to collect the data electronically and track it back to practitioners. That's really early, but we know that that's where we need to go, because, you know, many practitioners are actually doing this on paper and inputting it into their records, right? That's just the reality of, yes, of where we all are.

(Victoria): Me included. And I really, I guess, I want to emphasize to the committee that fact, right? Of, I'm a practicing rheumatologist, right? I've got a really busy academic practice. I do this with a piece of paper and text boxes. This is not a feasible measure. You guys all feel comfortable with it because you think it's a disease activity (unintelligible). This is not feasible, right? And I feel that really strongly.

All the other thing about the TB, that was feasible. This isn't. Right? Internists that are looking after RA will pull down on this measure. Many rheumatologists, probably more than 50%, who are really looking after their patients well, will fall down on this measure, because the EHR has not caught

up. Many of the EHR programs don't do the automatic calculation. You're waiting on a lab test to come back later. The patient is seen in the clinic, he's sent down to the lab, they don't wait to get their blood test done, they might come back three weeks later. Like, it's not feasible to do in the real world. And I really want to emphasize that about this measure.

Sheila Crawford: So we will -- real quick, just to kind of put the comments into place -- we'll circle back to that when we get to feasibility. Right now, does anyone have any comments or additional questions related to validity?

All right. So I think we can move to the vote.

(Asaba Denwulquafor): So the polls are now active for measure 2523, rheumatoid arthritis, assessment of disease activity. And we're currently voting on validity. Options are high, moderate, low, and insufficient.

We have all 20 votes in. Two committee members rated this as high, 15 rated it as moderate, two voted low, and one rated insufficient. So the measure passes on this criteria.

Sheila Crawford: Thank you. And now, moving on to feasibility. (John), is there anything you would like to say before we move to our…

(John): No. I, you know, I was just going to bring up that not - the date is not necessarily captured from a structured field, was the criticism, and then (Victoria)'s criticism of the complexity.

I think from reading what the developer -- and I'll let the developers speak, I won't speak for them -- was the fact that they used six tools and that some of them require lab and some do not, and that they did not require a specific tool

be used, that that was their rationale for why this would be feasible. But I'll let the developer address that.

(Lisa): So in terms of the number - in terms of general feasibility, so I think important to recognize that the way -- and I'm also a practicing rheumatologist at a busy academic center -- the way RA is treated requires an assessment of disease activity, right? We adhere to what's called (unintelligible) target, which means that we are supposed to be evaluating how active the rheumatoid arthritis is and directing our therapy towards it.

And so if you don't assess disease activity, you can't direct treatment appropriately. And there's a lot of debate about how to capture it, what's the best instrument, do you need a blood test, can you differentiate, you know, a patient who's experiencing pain from a non-rheumatoid arthritis condition? There are a lot of those technical challenges.

So when we selected those six, we did so in a systematic way, through a systematic process, with experts and literature, to support that those instruments were valid, reliable, you know, they generated precise estimates. Do they always entirely agree with each other? No. Are they very highly correlated with each other? Yes. And we did not feel that there was a best-in-class disease activity assessment tool.

And we also recognized that people have habits and they collected this activity in the way that they want to, and they may not want to change. So we tried to be flexible from that standpoint to allow them multiple ways to do this.

It is burdensome to collect disease activity, but it is an incredibly important part of treatment for disease, you know, for rheumatoid arthritis patients. And

the reality as well, you may not see this measure as feasible, you're still collecting disease activity on all your patients. And so - and you have an opportunity to not submit this measure to (MIPS), particularly as part of an academic institution where you've got a lot of - you're in a multi-specialty practice. I am. We don't submit rheumatology measures for (MIPS). We submit primary care measures. So it's primary care docs who are really shouldering the measurement burden with those kinds of measures.

So, you know, I think the (ACR thinks this is) really important. I think this is really important. It's a foundational piece for moving forward. We know that it's feasible in the registry, and we've worked with clinicians to make sure that it's feasible. Is it going to be feasible, you know, nationally across all EHRs? Not today. Do I hope we get there? Absolutely. Do I think the way to get there is by engaging patients and having patients collect some of this data? Absolutely. Would I like patients to be able to see their own data and track their own data? Absolutely. There are apps right now that do that for patients.

So I think the field is moving. I recognize that there's a lot of frustration with collecting this kind of data, but I think, you know, we've demonstrated that, in our model, it's feasible. We're working to understand how we can make it feasible going forward. I think that requires a lot of work with EHR vendors, which we're having those conversations, rheumatoid arthritis is not high on the list of EHR vendors wanting to change their EHRs to capture these kinds of things. But I do think that as the federal government moves towards interoperability, that we'll have some success from that standpoint as we move it towards being more systematically electronically captured.

Sheila Crawford: Thank you. (Devin)?

(Devin): So, a couple of comments. I agree with (Lisa) that this is an incredibly important measure. You're doing it. You know, your issue was with the feasibility of measurement.

And I guess my response to that is I too, and I was previously practicing rheumatology, doing it on a piece of paper, but I keyed in the results into the electronic health records. And I'm guessing you do the same thing.

(Victoria): We do too. I can guarantee they're not in a structured data field, they're (unintelligible) field.

(Devin): And they were not in a structured data field. But, you know, it's kind of, you know, a chicken and an egg deal here in that the EHR vendors, interestingly, they're not interested in RA, but if you turn on the television set, some of the common ads you're going to see are for Humira and Enbrel, for a 1% prevalent disease. And it's this exact disease, you know, activity which is driving the - or should be driving, you know, whether or not we're prescribing, which by the way are, you know, four of the five drugs that we use for RA and the top five, you know, most expensive drugs in the world.

So I think that, while I appreciate, you know, the concerns on the feasibility, if you make it a measure, then, you know, there's an incentive to core electronic health vendor - electronic health record vendors to, you know, put it in, you know, particularly given the kind of cost of these drugs that goes along with this measure. So, anyway.

So, agree, not a slam-dunk on feasibility right now, but I think it is doable. And also, as we kind of develop a more natural language processing, you know, capabilities, that thing that you kind of type in, that's not a (discrete) data field, will be something that we can pull out as well.

Sheila Crawford: (John)?

(John): So I'm not a practicing rheumatologist, but I'm on your side on this issue, as opposed to the last one we discussed. I can't imagine caring for this disease without measuring this parameter. You know, for years I've fought with pediatricians who didn't want to measure asthma severity as part of their routine care, and now it's a (ceiling) measure as best as I can tell, and we don't even have the MRs (unintelligible).

So this is, I think this is - and the other experience that's relevant, I was the core faculty for a - for-academic medical center collaborative on patient engagement and medication optimization in RA. And when those four institutions came in, it was - really, really difficult to do this, and they all routinely incorporated this into their practice through some simple systems redesign.

So, how they do it, that's not my forte, but they all did it, and everybody does it in asthma. So, I mean, how could you possibly care? This was shown in (Dartmouth) by (Gene Nelson) on caring for patients with spine disease and measuring their pain scores and all that kind of stuff 20 years ago when nobody was thinking it.

So I think it's just something that becomes a forcing issue, and if rheumatology community finds this really difficult, they got to get up and screen and storm your vendor. This is not a unique issue. I can't imagine why vendors wouldn't see that measuring functional status and disease activity is (cross-cutting) to many chronic diseases in an aging population. So I'm very strongly in support of this as feasible and important.

(Victoria): I'm going to respond to that. I actually, I want to be very clear on one other thing, I actually think that the physicians are measuring it, don't get me wrong. The physicians make decisions based on disease activity (unintelligible) no question about it. I completely agree that is how we manage RA.

That is not what I'm saying. I'm saying this measure as currently designed is measuring the ability of the EHR to abstract that data, right, and will falsely (ding) those providers whose EHR cannot meet that criteria, right? And I think there will be a lot of them. I think the reason you have good numbers in the (RISE) registry is that those people whose EHR doesn't automatically bring it out in structured data fields of presumably manually getting it into a field that will allow them to transfer that data to you, which is fine when the registry is small and is impossible when things become large, right?

And I think this is going to be really hard to implement. And I think that the danger with this is that you are going to - the people - the unintended consequence is that you will impact, you know, the people in certain parts of rural America who had been caring for patients with RA because there was no one else around, who now will be like, well, I'm not taking on those patients because I'm going to get dinged on this measure. Right?

And I think it will have unintended consequences. And I don't think the EHR has caught up with what the science, right? And I think that's been a dialogue that all of us have had multiple times with all of our institutions. Some of us who've been at multiple institutions have had it multiple times, right? It's been a constant dialogue, but I don't see us necessarily winning that battle in time enough to make this measure fair. Right? It's not measuring the quality of care delivered by that physician or by that group practice. It's measuring the ability of the EHR to pull the data.

Sheila Crawford: Ana?

Ana: So I can tell by Dr. (Shamagam)'s comments, which, you know, she obviously feels very strongly about, and I mean, I hear the point of the developer about we need to create an incentive for people to track these things in some sort of structured format. And if we do this, electronic health record vendors will then have some sort of an incentive or at least something on their checklist saying this is something we need to add to the next updated interface, whatever.

However, what's not clear to me, and I know I've raised this every one of these meetings that I've been a part of, I think, and I don't ever really feel like I get a straight answer, and there's probably a very good reason for that, but I'm going to raise it again.

What exactly is the implication if somebody fails to meet this? Is it really around, I mean, you mentioned it was a voluntary thing, does that mean that it's just - this is something we think is important, we're trying to create a process measure so that EHR vendors will be required to come up with some sort of structured data field?

Or is it, is somebody actually going to have a decrease in the amount of payment they're going to receive if they don't do this? Because I do think that the unintended consequences that Dr. (Shamagam) raised are important, particularly in rural areas where you don't have - I mean, it's very difficult to get a rheumatologist. It takes a lot of time. So if somebody who is, you know, also treating you for another - or another disease or another issue is brave enough to prescribe an RA drug, then, I don't want them to be penalized if their EHR isn't sufficient for that. I don't know.

Man: Yes. I think the - it's a tricky answer to your question, which is it all depends on how it's used, right? So that I think it's hard to say this will be the consequence of this, because it depends on what someone does with the measure once we endorse it and then use that endorsement as its justification for its use.

Woman: I think that's the concern, right? There are some of these quality measures that now are, I don't know, many years down the line, are being used to measure quality of care by physician or practice group, and impacts financial payments for that physician or practice group. Right?

I think there is some leeway right now as to a particular practice group might choose to report on certain measures, not on others. Right? And so those in group practices tend to have a little bit more leeway, might report as you do on primary care measures, as I do as well. But practices over, you know, what we don't know as we decide today is how this is going to be used five years from now, and I think that's the reality, right? And I would caution against adopting something that seems benign that has unintended consequences.

(Allen): Yes. And I think it's important to point out too that feasibility is not a must-pass criteria. So it is something that, when people look at how this committee voted, you can say that all the other areas were good, but we thought this is not very feasible at all, right? Which could caution people, but there's no guarantee that it does that. I think that's the important point.

(Scott)?

(Scott): So I don't treat rheumatoid arthritis, but as I get older I have lots of aches and pains.

So what I'm hearing is that it's only feasible in the people that use the registry, which is 30%. So what I'm hearing is that, in same percent of the people, it's not feasible. So, why should - it seems pretty straightforward that this measure and the way it stands right now is not feasible. Why should we vote for feasibility? Can the developers comment on that?

(Heather): Absolutely. So I want to address a couple of things. So, just to be clear, your last summary statement I think makes it clear, but, you know, we are measuring if you are collecting the data in a non-structured field. I want to be really clear right now that, no matter how you document your disease activity assessment in the, you know, in your notes, it is being captured. Okay? That there is - we are not missing data because - it's true for people in (RISE). I just want to be really clear, right? We are using - okay.

We are using natural language processing, I just want to be really clear. And we think that that is applicable to a large portion of rheumatologists. Do we think that we can get everybody in the United States who the rheumatologist onboarded in the next year? No. Could we do it over the next five years? Probably. Right? If people were engaged, right? The cost of registry to ACR members is zero.

So this is a no-cost solution, and I would say upwards of 95% of rheumatologists in this country are members of (RISE). So this is a - I'm getting a headshake, but a large majority of rheumatologists are members in (RISE), and if (Tracy) has the number off the top of her head, or (Amy) has a number off the top, they can tell us. But this is a free service to rheumatologists.

So I will argue that it is, and, you know, you made the point, (Allen), that it's not a must-pass criteria, but it is feasible and that we are moving forward in

making it increasingly feasible over time. And I do think that, if you don't endorse a measure like this going forward because it gets hung up on the fact that there is a burden associated with a clinician somehow documenting in their chart that they performed an essential assessment of their patient, then (Cathy)'s point is exactly right, which is the EHR vendors are going to say, "It's not NQF endorsed." Now, this does not have to be NQF endorsed to stay in MIPS, but it helps to have the vetting process that goes through here.

So I think it's important to recognize that, for the people who are in this registry right now, it is very feasible. And there is a question of feasibility for another proportion of rheumatologists in this country, and I think it's a really important message that this committee sends to all developers and to Medicare and to everyone about, you know, EHR vendors, about moving us forward with having some flexibility. I would actually much prefer that that's about the patient access to this data. Right?

Like, we can talk about what's burdensome for us, but I would actually think that the conversation is much more about, how do we make it less burdensome for our patients? Right? How do we make this data available to them, so that they can track their own disease activities?

So I think there are lots of issues in terms of implementing this over time. And I will also reflect that, last I checked, it was every three years that measures come back. So this will not sit out there for five years without a committee reevaluating it. It is going to come back in three for a reevaluation of feasibility and where we are and what the EHR vendors are doing in terms of keeping us, moving us towards being more feasible. Thanks.

Sheila Crawford: Thank you. (John)?

(John): I was just going to add a comment and ask (Victoria)'s opinion, it's only a 50% measure to - meaning you'd only have to do it at 50% of the time. So, does that make it…

(Victoria): Fifty percent of visits per patient has been generally every six to eight weeks, right? So we would see them six times a year. So, I mean, I think if you're doing it, you're doing it. If you're measuring it, you're measuring. Unquestionably, right?

I actually do measure it every visit, it's simply that I know it would be impossible to pull that data. It's in a text field, right, that is part of the physical exam, right? And I just look at that as that's how I do it, as I've kind of, you know, developed my practice. Other people will have other work-arounds, right?

I think what's important is I don't think that the cutoff of 50% necessarily helps you. I think that if people are doing it, they figured out a way to make it measurable. My concern would be the dinging of the people who are actually providing the quality care and get dinged on this measure because they're not able to extract the data in a way that meets the measure.

Sure, that is a reason to advocate for better EHR. I feel like that's a battle we've been fighting for years and not won as rheumatologists. I don't know. I think you have - your altruism bucket is fuller than mine, maybe.

Woman: I want to ask the developer, when you use your natural language process, are you attaching that to a (SnowMed) code that goes over to an ICD 10, so that it could be a process that could be easily and feasibly applied?

Woman: I don't think we can answer that right now, but we are flagging it in some way. Whether we're mapping it to (SnowMed) codes, I don't think so. But we are mapping it (unintelligible) great. We're mapping it to the data elements that are in the measure. So it could eventually be matched to a (SnowMed) code. Your point is excellent. Yes, thank you.

Sheila Crawford: (Heather).

(Heather): Just one comment (unintelligible) feasibility discussion. There are practices that I know of who used, you know, this measure as just an internal quality assessment, you know, QI, you know. So you're measuring it. So it's feasible in your practice.

Woman: (Unintelligible)…

Sheila Crawford: Can you turn on your microphone just to make sure the folks on the…

Woman: So we are measuring it, I think, and we are assessing it internally. But I know that we wouldn't meet the measure as documented and - for this quality measure, right? We wouldn't meet it because it's not pullable. It's a manual.

(Heather): No, no, when I say - meaning that you're using it as a standard of care or that you're actually measuring how often this is assessed within your population.

Woman: Yes, but we're pulling that data because they're human beings (unintelligible). That's not feasible - that's not scalable. That's my concern. Because I can't, you know, that's fine as a quality improvement project, but it's not fine in real life, right?

(Heather): Right. So the practices that are functioning exactly like you who are being measured in (RISE), they are, you know, we are pulling the data from their notes and we're, you know, we have an understanding, clinicians work with the EHR vendor and they map the natural language processing to the note and they pull it out of the unstructured note. So it is indeed feasible, your question of scalability is an interesting one, right? It really depends, right? Scalability depends on your timeline and your commitment and the institution's commitment. It's really not more than that.

Woman: Correct. And just to clarify a point made earlier, I guess, I think what you were trying to say earlier was 95% of practicing rheumatologists in the country are members of ACR. And anyone who is a member of ACR could choose to enroll in RISE. However, in order to enroll in RISE, you would then have to get buy-in of your organization to agree to transfer data, right, which is obviously then under the auspices of a different kind of legal process of data sharing. And I think that, I think is why only 30% of practicing rheumatologists actually transfer (data), right?

(Heather): I think one of the - we know that some of the restriction was people have EHRs that, up until now, were challenging. I named Epic because it's a huge proportion of academic centers and we had enormous numbers of academic centers who said, "I can't move this forward because Epic doesn't let us move data." And we have worked, and I'm happy to let (Tracy) from the ACR, (Tracy Johansen), address how we work with EHR vendors to ensure that we're trying to move feasibility forward. Certainly the Epic I think conversation is one that was very painful but has led to success and we, you know, now are onboarding our first Epic practice. I don't know if, (Tracy), if you want to just add any comments about that.

(Tracy Johansen): Yes. I would just note that with the recent development with Epic, we have an influx of academic centers who've expressed interest in joining the RISE registry, and we had to actually put some of them on pause because we want to do some pilot onboarding of just a handful of centers so that we can then take lessons learned and implement that across the board with any academic centers moving forward. So there is always the, you know, the time and investment it takes to make the connection, and especially now that we're able to access that data that is collected in Epic, that's going to significantly improve our numbers and diversity of the data in the RISE registry.

I would also add that the ACR is actually working directly with some of these EHR vendors to implement even more structured fields specific to rheumatology. A number of our volunteers spent the time putting together templates for care of rheumatology patients, specifically around RA, psoriatic arthritis and lupus, in order to make sure that there's consistency in the data that's being collected and how that data is being collected. And we are now working on taking that information to these electronic health record vendors and getting this information implemented into structured fields, so that, you know, we increase the likelihood and ability of us being able to scale this and onboard practices much more ably and much more quickly? And one of the, obviously, one of the most important elements in that template is the ability to collect these disease activity scores.

Sheila Crawford: Thank you. Any final questions or comments about feasibility?

Okay, we can go ahead and move to the vote.

(Asaba Denwulquafor): So we're currently voting on feasibility for measure 2523, rheumatoid arthritis, assessment of disease activity. The options are high, moderate, low, and insufficient.

And polling is now closed. Ten committee members voted moderate and 10 committee members voted low.

Woman: So that's consensus not reached, 50-50. But since this is not a must-test criteria, we'll continue on. And we will still vote on overall recommendation for endorsement.

Sheila Crawford: Okay. Use and usability. (John)?

(John): It was previously (an e-measure) as a part of PQRS. It is now, as discussed in - as a part of the RISE registry, and apparently it's being implemented into MIPS, so that they'll be, by 2020, be able to pull data. And its results have been publicly reported. They do give feedback to the participants in RISE as to how they're doing. And it was a recommended pass on use by NQF staff.

And relative to comments, there was again the concerned voice about RISE representing only 30% of practicing rheumatologists.

Sheila Crawford: Thank you. Any additional comments on use and usability? All right, I think we can move to the vote.

(Asaba Denwulquafor): Poll is now open for the use criteria measure 2523. The options are pass and no pass. We're waiting for one vote. There we go. Fifteen committee members voted pass, and five committee members voted no pass on criteria of use for measures 2523.

Polling is now active for the usability criteria. The options are high, moderate, low, and insufficient. We're waiting on two votes. We're waiting on one more vote. So, for measure 2523 usability criteria, one committee

member voted high, 14 voted moderate, and five voted low. So, 2523 passes the use and usability criteria.

Sheila Crawford: Thank you. I don't believe there were any related or competing measures, correct? So we can move on to overall suitability.

(Asaba Denwulquafor): Overall suitability for endorsement, the options are yes or no, and this is for measure 2523, rheumatoid arthritis, assessment of disease activity. Overall suitability of the 20 committee members who voted. Fifteen voted yes and five committee members voted no. Measure 2523, rheumatoid arthritis, assessment of disease activity passes on this criteria.

Sheila Crawford: Thank you. So we've been going about three hours, so I think we're going to take our break. So if folks would like to take a break and come back by 11:00, is that good, 11 o'clock? Awesome.

Operator: The conference has been unmuted.

(Lisa): Hi, can you guys hear me?

Woman: (Lisa), is that you?

(Lisa): Yes, it is.

Woman: Great. Okay, we can hear you. So we're all set.

(Lisa): Okay. Thank you.

Woman: Thank you.

(Lisa): Okay, everyone, we're going to get started, if everyone can get back to their seats.

Sheila Crawford: So as folks are going ahead and taking their seats, our next measure is 2525, rheumatoid arthritis, disease modifying anti-rheumatic drug therapy. And our lead discussant, William Curry. If you could give us a brief overview of the measure and then lead us directly into specific comments on evidence.

Bill Curry: Sure. This is individuals who are 18 and older with a diagnosis of rheumatoid arthritis, who were seen with two or more face-to-face encounters, with the same clinician during the measurement period. That's the numerator statement - I'm sorry. That's the denominator statement.

Woman: Mic please.

Bill Curry: Mic is on. Sorry.

Denominator statement is 18 years of age and older, with the diagnosis of RA, and had two or more face-to-face encounters with the same clinician during the measurement period. The numerator is that they received a prescription for DMARD. I will state that this is a little different than the HEDIS parameter that we have in primary care been held accountable for, where it's based on claims data. So that's the responsibility of the patient to be adherent to our negotiated recommendations. So this is not based on whether the patient is adherent to medication use.

I believe that the evidence is moderate as described by the committee. There was evidence-based guideline that the ACR has posted that has systematic review. There's a logic model that I think is, I think, very appropriate to get to the treat-to-target that we've talked about earlier to achieve remission or to

have low disease activity through the course of their illness. So their guideline was published in 2015.

The gaps in the evidence, so, actually, the performance has been pretty high in the RISE registry, with over 90% adherence in the inner quartile range is actually very narrow at 6.42, which would beg the question, are we at a ceiling for performance with this? The HEDIS parameter recently was reviewed for the same reason as that they had very high performance and some thought that there was a ceiling with the claims-based measurement of DMARD prescription.

There were some other issues with the HEDIS parameter that was bringing it to discussion for retirement, but the question I think is, are we doing really well with this in terms of rheumatologic care?

The data is from the RISE registry, as we've heard before. There is no disparity or (SDS) data that's collected as part of that, but there are data that would suggest that, especially for Medicare Advantage plans, that there are some racial disparities, low income disparities, age disparities, and also regional disparities in this care of patients, so, but not captured within this measure or in the RISE registry. I would recommend that this is a moderate evidence.

Sheila Crawford: Thank you. Any additional comments from the committee related to evidence?

Okay, I think we can move to the vote.

(Asaba Denwulquafor): So the polls are now open for measure 2525, rheumatoid arthritis, disease modifying anti-rheumatic drug therapy. The options for evidence are moderate, low, and insufficient.

We're waiting on two votes.

So we have 20 votes in. And all 20 committee members voted moderate on the evidence criteria. So the measure passes this criteria.

Woman: Can I ask a quick question of the developer? In regards the exclusion criteria of patients with inactive rheumatoid arthritis, how is that captured, evaluated?

Woman: That's captured by the physician indicating it. We can either, and that depends again on the different practice. So we do capture disease activity assessment, so they can indicate that if the patient's in remission, that there's a marker for that. Or we can capture it in their notes if they have an indication. But they work with the vendor for the registry to ensure that inactive disease is captured in the data to exclude those from the denominator.

Woman: So, to clarify, and I guess this is linguistics, but there would be a difference, correct, between an inactive problem of rheumatoid arthritis versus rheumatoid arthritis disease in remission, and is that distinguished?

Woman: I believe it is distinguished, that one is through coding an inactive, and in our experience, it's - coding is actually challenging. So they are distinguished in our data.

Woman: So I ask only because again I do think that this might merit some thought back to the development committee, again with that question of scalability and what happens to the internists who are seeing patients with RA. Obviously it's

an active problem, right? It's not an inactive problem when that person's (shopping). They may only not be in remission, they may only not want to be measured in terms of that performance on that active problem. And just if the committee, as they continue to develop and refine this measure, could put some thought into that, I think it might help the larger community.

Woman: Great feedback. Thank you.

Bill Curry: That was a complaint in the HEDIs parameter, that there was no way to take those patients out who were low activity or in remission, that they did DMARD. Okay. For reliability, the medication…

Sheila Crawford: We haven't - we have to…

Bill Curry: I'm sorry.

Sheila Crawford: …do performance gap real quick.

Man: I appreciate the (unintelligible).

Sheila Crawford: Does the committee have any additional comments related to performance gap?

Oh, I'm sorry. (Lindsey).

(Lindsey Bachford): So I have a question, and maybe it's best addressed by the developer. But given that the performance is so high, you know, that the mean of 90%, and in talking about the disparities, it talks specifically about the decline in use with age. And so I'm curious, for the 10% that's not, you know, that is in the gap, is it because it's maybe a decision by the clinician that their age is not

indicated, or is there thought for why that gap still exists even though it is relatively small?

Woman: I think there are a couple of issues in your question, one of which is the size of the gap. And yes, it's narrow. We do have - actually, this is the highest-performing measure that we have on - in our registry. But as we, you know, it's interesting, when you onboard practices, they learn rapidly. And so some of it is learning in the backend of the data that they need to be mapping, and there's an iterative process for mapping the data for accuracy. And some of it is just learning that they're not doing things that they thought they were doing.

Or one physician in the practice suddenly stands out as being very different than the other physicians in the practice. So there's a lot of learning that goes on, that we think is still valuable, particularly as we have additional practices joining RISE all the time. And this is a critical measure I think for patient care, and others have highlighted the concern about disparities.

And then the second part of your question, can you just help me - I think was a little bit less about gap and more about sort of data availability?

(Lindsey Bachford): So I guess I wonder, with the gap being so small, I wonder if there really is potential to close it more, or if a clinician would choose to just say "inactive," you know, inactive disease to meet the measure, because they know it's an exclusion factor. I'm questioning whether that gap still warrants a performance measure, in my mind.

Woman: Great. And you also asked about aging. So I think one question is, is there a possibility of gaming? Yes. We work very hard to design measures that aren't gameable. As we're moving towards actually measuring the outcome of disease activity, not just whether or not you're capturing disease activity and

how you're treating it with your disease modifying agents, I think we'll have much better data understanding what's truly inactive disease via disease activity assessment. So I think there'll be less chance for gaming over time.

In terms of aging of the population, there are patient preferences that change based on different demographic groups, but the fundamental phenomena of inflammation that drives the need for treatment exists. So, you know, we still think that there's an important, you know, ideally, you may never get this to 100%, but I think even with an, you know, average of 90%, there's still a meaningful gap, particularly as we are bringing new practices on all the time, and as we've talked about earlier, we're trying to understand the role of disparities in this area more.

Sheila Crawford: (Victoria).

(Victoria): I had a bunch of other questions for the measure developer. Sorry. Again, some language clarification on some of the exclusions. One was a point that I wonder whether you wanted to change your patients with the diagnosis of HIV to patients with an infection that precludes treatment, that would then capture your atypical micro-bacteria patients that you don't actively want to treat today but you might in a couple of years, your patients have Hep B where you're addressing the Hep B for like a couple of different situations and scenarios and not just HIV, and it would say, I personally care for some patients with HIV, where we are actively treating their RA in addition to actively treating their HIV. So I'm not sure HIV per se should be necessarily an exclusion.

My second question was in regards what you're actually measuring. So this I guess gets back to some of our unintended consequences with some of the things that we think are appropriate and are going to change behavior. And

are you in this measure measuring prescription being written? Are you measuring prescription being taken by patients? Are you measuring the decision point at which the physician provider or providers accesses disease activity and institutes an action plan, right? They were assessed for appropriateness of DMARD. Yes, no. And was the DMARD given? Yes, no. Right?

I guess, could you explain a little more how this measure is going to work? Because what it seems when you - when I read it, was the patients receive the DMARD, i.e. a prescription was written. And I can think of plenty of people where a prescription gets written but not taken. Can you just give us a bit more info?

Woman: Great point. So, first of all, thank you for the comments on, you know, modifying or expanding or changing the denominator exclusion criteria, those are great feedback.

The measure assesses the providers' actions of your placing, you know, your prescribing or placing the patient on a disease modifying. It does not capture whether or not that patient picks that up. So there is definitely a patient compliance issue, absolutely. We are not like the HEDIS measure checking as to whether or not that prescription was in fact field. So, absolutely there's a potential for overstating the number of patients that are on a disease-modifying agent, and I certainly think that that's another reason for indicating performance gap. It's also the reason why we're trying to measure outcomes and not just process measures.

Process measures are helpful to tell us what we need to do when we don't have the outcomes that we want, but ultimately we want to understand how this relates to their actual disease activity. So we're working on that.

(Victoria): Follow-up question to that, and so, is that data being pulled from a field in EHR prescription field, that something (unintelligible) or DMARDs, so, you know, (unintelligible) check that box, or would it only be methotrexate, right, one of the more potent DMARDs. And then secondly, how are you capturing, so for my practice certainly, infusion medications, IV infusion medications, it's a different ordering process, it doesn't go through the same prescription database actually, and so it would be again very challenging I think to make sure that that data wasn't being missed but your Remicade patients (unintelligible).

Woman: So as with the other measure, we are capturing it in a number of different ways, right? So in an electronic health record where you have e-prescribing and it's all integrated, it pulls from the e-prescription.

But there are a lot of, particularly outpatient EMRs, where that's not integrated. We can either integrate it for you, like some people have e-prescribing in a totally separate health record, that's separate from where they document their medical notes. We combine those at the patient level. In others we're pulling from notes, you know. So it's a range. And I think it's a parallel conversation to the conversation we had over the last measure, in terms of feasibility, which I know is not what we're talking about right now.

Woman: Sorry. Then I guess…

((Crosstalk))

(Tracy Johansen): This is (Tracy Johansen) from the ACR. I just want to one other note here. There's also medication data being pulled from the medication table. So, every time a provider sits down and goes through the current medication list

with their patient, that information is also being collected in the (res) registry as well. And for the majority of our practices, we are able to see - excuse me - we're able to see - they're able to note whether or not the prescription was written by the provider or by an outside provider.

We also have, you know, the ability to look at the NDC codes, RX norm codes, HCPCS codes that come out of these tables as well.

Woman: Could you speak to the question of infusion medication? So, and I'm particularly thinking of the scenario in which private practice rheumatologists perhaps uses a home infusion company to deliver Remicade.

(Tracy Johansen): So if a medication - you're saying they have a home infusion company but that med's - prescription is still written by the provider, correct?

Woman: But it doesn't get sent to a pharmacy. It won't necessarily be written in the EHR and it doesn't get sent to a pharmacy, the same pharmacy as like your CVS or Caremark or whatever.

Woman: But it's captured in the medical - medication reconciliation table.

Woman: It may or may not be.

Woman: That's the point of the medication…

((Crosstalk))

(Tracy Johansen): Is there reason why that information would not be put into a medication table or captured during medication reconciliation?

Woman: Yes, because the person isn't writing that prescription in that way. I can think of a scenario where that does happen, and that is because the prescription, the actual physical prescription, is written separately from the EHR.

Woman: So you're talking about using an e-prescription module in order to write that order?

Woman: No, no, no. So I think this is probably geographically dependent, but in certain places in the country, there will be private infusion companies that either are running a freestanding infusion center, where the med is being infused, or are literally going through the patient's home and infusing the medication at that patient's home. But it's usually written on paper. It's usually - the actual order itself is usually a piece of paper (unintelligible) in the chart. And I ask just from experience I could see that scenario happening.

Woman: So I mean, I think similar to the last measure, in situations where you have unique and unusual prescription practices, that's going to be documented somewhere in the medical record that the patient is on this medication, and we capture that. We may not be capturing it in a standardized field for that group or that individual practitioner, but we're capturing that information.

Woman: And I would just add that part of the, as we mentioned earlier, that handholding with onboarding with her practices is ensuring that we are giving the information where they're capturing it. So, during the review, they're seeing that some of the medications that might be kept, prescribed in that manner, are not being captured, they're able to work with our technology center and get that area where they're making note of that, outside of the medication table or the e-prescription module, mapped for their data, included in measure calculation.

Sheila Crawford: Thank you. Ana?

Ana: One question I have is, and this is sort of a concern that I have with many of these measures, is, is it appropriate to hold a physician or a hospital accountable for something when the insurance company is making it nearly impossible for the patient to access it at an affordable level? I mean, you know, I have acquaintance who have Type 1 diabetes and rheumatoid arthritis, she's on Medicaid, and is ratcheting insulin because, you know, insulins are expensive and copays are ridiculous, and I mean, it's, I don't know, I mean, I've had certain instances in which my physician was excellent, would prescribe certain medications, and then the insurance company would deny it. So I'm just concerned that we're holding physicians accountable for decisions that they have, you know, no ability to influence a very limited ability.

Man: Maybe I could speak.

Sheila Crawford: I also want to make sure we stick on performance gap. So, some of the comments, like, they're kind of bleeding into the other sections and I want to like keep them in that section. So, is your comment related to performance gap? Okay, go ahead.

Man: So, regarding these both comments, I'm one of the medical directors of our electronic system, and medication reconciliation is supposed to actually alleviate and solve both of your issues. So, even if it's on paper the next encounter, it should have been reconciled, and vice versa. If the insurance company never approved it, it should have been reconciled as well. So I think in both cases it would have took care of that situation.

(Victoria): Well, except it would have taken it off, right? So the person, the provider was trying to order the medication and the insurance said no, and so it then gets

taken off to a visit two. Isn't it two visits? Visit two, now the prescription is no longer there. I don't know. I don't know how you adjudicate that. I think it's a great question.

Woman: So if it gets removed and discontinued, it will get pulled out. I think the other question was really about implementation of the measure in terms of what's the impact of measuring physicians on prescribing when there are access issues to medications. And, you know, we can maybe address that during usability.

Woman: Okay.

Sheila Crawford: (Unintelligible).

Woman: Just in extension to (Victoria)'s comment about exclusions, that I think one more thing that's important to add is active malignancy, because a lot of these patients can have (flare) of their disease and/or have rheumatoid arthritis occur as, you know, around the time of malignancy, and if they're being actively treated for that, you may - it may not be feasible to start a DMARD, even though the drugs they're on may not be controlling their disease. So I think that needs to be taken into account.

And part of this is also, you know, this was brought up many times about patient preferences, does the registry capture a patient's preference of not wanting a DMARD? And how does that then act against a physician's responsibility to provide it?

Woman: So I will ask (Tracy) to capture whether or not we have patient refusal as an exclusion. I don't believe that we do. That's a really game-able issue, and it's a really important thing going forward to talk about. I will say that the choice

of a disease-modifying agent in rheumatoid arthritis is very - very much requires shared decision-making, and there is a huge range now of disease-modifying drugs, some of which are really expensive, and some of which aren't, and that there are almost always options for patients to work around comorbidities, malignancy.

Your points are all really well taken, but rheumatologists actually do have a lot of tools at their disposal in terms of potential treatment that would meet this measure, even if it may not be fully adequate for treatment of the disease in the optimal sense, even for patients who are of limited access. I think it's a great issue in terms of whether or not we capture patient refusal and I'll defer to (Tracy) on whether or not we do capture that.

(Tracy Johansen): That is not currently included in our exclusions.

Man: (Jamie). Oh.

(Victoria): So in development was there a reason that you didn't include it? Was it just the concern that people were going to game the system?

Woman: Yes, so people ask us, I mean, this committee whenever we brought it before, there was a question of can't you - can't there be an out for providers to say patient's non-compliant? And we just really, over and over and over again, and speaking with the patients, we just felt like that was a risky, maybe slippery slope to go down, and for all of these measures. I think it's important to keep revisiting it and revisiting the impact of it. But we were uncomfortable giving clinicians that out for this measure.

Woman: I think gaming the system can go both ways, right? Because if you're going to put a measure in place saying that you have to prescribe the D-marc, and if the

patient does not want to be on it, you can easily game the system and put a prescription in the records and the patient won't pick it up. So I don't know if gaming really works only one way in terms of refusal.

Woman: That's a great point. There absolutely can be gaming in either direction. You know, these measures are imperfect, that's why we have a suite of lots of different measures and not one in and of itself in isolation. I think our hope is that we're fostering a conversation to find the best medication for a patient, and that requires a lot of conversation with the patient, and this certainly encourages, if the patient's first response is I don't want to take this medication, and you as a clinician believe that there is a really strong medical reason for them taking that medication, you need to understand why they're refusing it.

Because they may be refusing it for reasons that are work-around-able. Right? It's an infusion and it's costly, all the things that we've been talking about today, and we as practitioners are required to address those concerns and manage them and ensure that the patient's getting the appropriate care that they need.

Woman: I completely understand that and we all as physicians want the best for our patients and we will in that setting do the right consort, but as a measure, I think it's important to understand in this particular measure that it reaches 100%, then something's wrong, because it's not one of those measures where you can get 100% buy-in from both the patient and the physician.

Woman: I was just about to say the same thing.

(Bill): How would attribution of responsibility for this measure be made on an individual level? Is it just the specialist, the rheumatologist, or is the primary care doctor, internist, responsible for having this taken care of?

Woman: Currently it's the prescribing physician, so…

(Bill): So but if there's no prescribing physician, then you don't - in other words, if it hasn't been prescribed, it may not be the prescribing physician. You may not be able to attribute.

Woman: So you get - sorry, you get into the denominator for the measure with two encounters with the same practitioner. So you need to - that addresses some of the concerns about the primary care doctor seeing the rheumatoid arthritis patient in this measure. And then you capture the - whether or not the D-mart has been prescribed, whether it's through the medication reconciliation table, and your concern is that at this point it's not being attributed to primary care doctors.

It's only being attributed to rheumatologists. I don't think there's a risk of misattribution at this point, but it's certainly a point we can consider going forward with the other issues about outpatient primary care providers.

(Bill): I have to think that there are a good number of non-rheumatologists that actually prescribe these drugs, and what - yes, so what happens in those situations? They're just not on the radar?

Woman: Right now, they're not on the radar. It's a great point.

(Bill): So you may be - actually may be underestimating the gap.

Woman: Well, I know we're underestimating the gap, because as we've talked about before, we're not measuring all the rheumatologists in the country, and we're even if we were, we're not measuring individuals outside of rheumatology, but you know, we're starting small, I guess is the response.

(Bill): Thank you.

(Victoria): I think I had two - kind of two final follow-up points on that, but yes, the question of gap, right? You almost don't expect it to be zero, right? There should always be a performance gap, because there are going to be some non-compliant patients. I had a question about the D-mart, as in it says a D-mart. It doesn't say what (unintelligible) of D-mart are being considered to be. And I guess my question is Prednisone? Does it count? Does it not count? Because everyone would agree that prednisone isn't the perfect lone therapy, sorry.

(Bill): Yes, there is a listing in the specification of D-mart.

(Victoria): So it's in the Cognos.

(Bill): And prednisone is not one of the D-marts.

(Victoria): Okay, good. And then secondly do you think that there is an unintended consequence of over-prescribing as we look at this measure?

Woman: What do you mean by that?

(Victoria): I mean patients who don't need a D-mart, people who are in remission.

Woman: So given the documenting inactive disease as an exclusion for the measure, I think that risk is low, but again all the more reason to move us towards other measures over time.

(Victoria): I would again clarify that word then. I think we have to have disease remission as opposed to inactive, because I guess I'd question what you mean by inactive, as in inactive on that problem list, they didn't really have it? They took the problem off, or they have it but it's in remission? Are they in remission on therapy? I just would clarify the language.

Woman: Great, great suggestion, thank you.

Man: Any other comments on gap? Okay, we can move to the vote.

Woman: So (unintelligible) active for performance gap for measure 2525, rheumatoid arthritis, disease modifying anti-rheumatic drug therapy. The options are high, moderate, low, and insufficient. Waiting for two more votes. The results are now in, for the 20 committee members who voted for the performance gap, all 20 voted moderate on that criteria, so the measure passes this criteria.

Man: Right, moving on to scientific acceptability, reliability and validity, (Bill)?

(Bill): The agents are listed in the specifications, and in testing, using signal to noise ratio, the mean reliability was 90%, so very high. It was not broken down in the testing that I remember by individual provider, or separated by individual provider.

Woman: Planned recipients, the highest rating achievable. Find out more.

(Bill): So I think the reliability is high. Moving to the validity, they tested two large systems, one a Serner shop and one an Epic shop. One had a more rural bent and one was a combination of urban and rural, and they looked at chart review versus registry data. There was some discrepancy in the diagnosis of rheumatoid arthritis between the two, and it was thought to be coding issue by providers at the visits.

And there's ongoing work in the rise registry with annual data validation and with their onboarding process to do ongoing chart reviews to help to reduce that. Face validity was accomplished through what seemed like an arduous system with experts and other groups including payors and patients and the face validity was high at nine. When they looked at CAPA scores, the CAPA coefficient was .67 overall.

And I think that's pretty good in terms of reliability and for the validity as well. They also looked at the exceptions, and they found that the exceptions in the rise registry were actually very low for HIV, pregnancy, and/or those with no evidence of - with no active disease. Interesting, they stated that there's no missing data, and yet the literature would state that maybe EHR data in terms of prescriptions, 15% omissions or things on the active list that shouldn't be there.

So there must have been some good data scrubbing for that to be no missing data, but I think that will potentially be an issue if this measure goes outside of the rise registry. But I think for reliability and validity, I think that it's accessible.

Man: Thank you. Any committee comments on reliability?

(Bill): I'd just point out that she's posted the individual level testing here, again the minimum threshold of 10, very, very high reliability.

Man: (Jamie), did you have something?

(Jamie): No, I (unintelligible).

Man: I think we can go ahead and move to the vote.

Woman: Pollings are active for rheumatoid arthritis disease modifying remark, anti-rheumatic drug therapy 2525, for reliability, the options were high, moderate, low, and insufficient.

(Victoria): Can I ask another question of the developer? What are the plans to update the list of eligible D-marts on a regular basis?

(Starlyn): So we do this annually, and (Tracy) can add the timing, but it's done on an annual basis. I think we just underwent it. I don't know if we've actually closed the ability to update for this upcoming cycle yet.

(Tracy Johansen): Yes, we do review the list of drugs annually along with the codes associated with those drugs, and that's usually done around the beginning of the year, so in about another six months we'll get that - we'll do another round of that.

(Victoria): And how - do you mind me asking, what is the stakeholder input into that process, the rheumatologists' input?

(Tracy Johansen): So we don't solicit formally, but we get feedback. So just in the last cycle we had feedback from I think two or three rheumatologists out in practice who are using the measure who make an observation based on their practice or

their knowledge of the literature and updates get made on that basis. So we're open to feedback at any time. We don't formally solicit input on an annual basis from the membership, although we could.

Woman: So those are the results for reliability criteria, 7 committee members voted high, 12 voted moderate, and one committee member voted low, so measure 2525 passes on the reliability criteria.

Man: Thank you. Any comments related to validity? Okay, we can move to the vote.

Woman: Measure 2525 validity, options are high, moderate, low, and insufficient. Just waiting on one more member. The results for validity on measure 2525, rheumatoid arthritis disease modifying anti-rheumatic joint therapy, three committee members put as high, 16 committee members choose moderate, and one committee member voted low. So the measure passes for this criteria.

Man: Thank you. Moving on to feasibility.

(Bill): So the data field that the measure would be capturing are discreet data fields, visits, ICD10 codes for the visits, the prescriptions for the D-mart, I think they would all be discreet data fields. There may be some absences as we talked about. I think there are actually, in my MR if somebody is getting a prescription prescribed by somebody out of the system, I can add it by history, so it still shows up on the med list. But I think that - for the most part I think that this is a feasible measure.

Man: Any additional comments related to feasibility? Okay, we can move to the vote.

Woman: The poll is now active for feasibility. The options are high, moderate, low, and insufficient. We're voting on measure 2525, disease modifying anti-rheumatic drug therapy for rheumatoid arthritis. Two committee members voted high, and 18 committee members voted moderate for 2525, rheumatoid arthritis, disease modifying anti-rheumatic drug therapy so the measure passes on this critiera.

Man: Thank you, use and usability.

(Bill): So currently I don't see that there are any public reporting programs or payment programs that this measure is being used. It used to be a PQRS measure. It's not been moved to MIPS. There's work being done within the rise registry for recognition and for quality improvement activities. So I think that within the uses of rise, I think it's very usable, as this measure perhaps steps out of the rise registry use.

I think it will be something that will be relatively easy to have systems put into place. The question will be, will it remain in scripts being filled - I'm sorry, the scripts being written versus scripts being filled, switch more to a claims based process, but as it's written, I think that it would be just the prescriptions within the record, so I think that it's usable.

Man: Thank you. Any additional comments related to use or usability? Okay, we'll move to the vote. Vote.

Woman: Poll is active on the use criteria for 2525 for rheumatoid arthritis, disease modifying anti-rheumatic drug therapy. The options are pass and no pass. Of all 20 committee members who voted, we have 20 members who choose pass on the use criteria. The measure passes on this criteria. We're now going to move to the usability slide. The polling is active for usability. The options are

high, moderate, low, and insufficient. On the usability criteria for measure 2525, we have two votes for high, 17 committee members choose moderate, and one committee member voted low. The measure passes for this criteria.

Man: Thank you. And now onto the related and competing measures.

(Bill): There is one related measure, and that was the HEATUS measure from NCQA that I had mentioned earlier. It's a claim-based D-mart account. There was no exclusion for inactive rheumatoid arthritis. It's a plan based measure. It was being reviewed for retirement because of this issue with the ceiling effect, the issue with how to handle inactive disease.

And I don't - the last document I could find from NCQA was dated the end of March this year, and I don't know what their final decision was, but I think there are differences between the two measures. They also found though that it was set up a pretty successful rate of completion, of success.

Man: Got it.

(Starlyn): So I think a big difference between the two measures is that the NCQA measure is a plan level measure, right? And so they measure it to commercial, Medicare, and Medicaid, and there actually was a continued performance gap in the Medicaid population, though less so in the other populations.

Man: Is there a particular step we need to take in relationship to this?

Man: There is not competing.

Man: Got it.

(Starlyn): I will note, we worked with NCQA to harmonize to the extent that the data allowed, but (unintelligible) matters.

Man: So with that, I think we can move onto overall suitability. Do we need to review the public comment? Just review it. Yes.

Man: Reading it?

Man: I'm sorry, this is the first time I've had a public comment come up. It's sort of the measure process, but I just need to check about process on it. Do we need to discuss it, just acknowledge it?

Man: You may discuss it if you wish, but yes. Minimally we should acknowledge the competence thing. Right.

Man: Yes, there was a public comment related to the use of brand name drugs. I'd say on the patient side I appreciate seeing brand name drugs, because that's the only way it's ever spoken about to us, but I can appreciate other people having different viewpoints.

(Starlyn): It's a recommendation that we add some anti-clinical drug lists to our mapping protocols, and we'll definitely take this under review. Assuming - if it's feasible, we'll definitely do it. I don't know much about this, if it's…

(Bill): We have a plethora of drugs going to generic at some point, which actually makes a lot of sense.

Man: Any other comments. All right, I think we can move to overall suitability for endorsement.

Woman: Voting is current for overall suitability for endorsement, measure 2525, rheumatoid arthritis, disease modifying anti-rheumatic drug therapy. The options are yes or no, and all 20 committee members have voted, and we have 20 yes votes for this measure, so the measure passes this criteria.

Man: So we're powwowing on the time schedule over here, just one second.

Woman: Thank you very much.

Man: Okay. Oh, before we say it, thank you so much. We appreciated you had a lot of questions, and we appreciate your participation as well as those on the phone today. So what we're going to do now is we're going to go ahead and move through the PQA measure and then we'll break for lunch. Sound good? That way we can kind of - you looked really sad there when I said that.

And then also before we begin these two discussions I just wanted to mention that I spoke with NQF during the break, and there's not a conflict, but I have done some collaborative work with PQA and I'm about to with PCPI on patient centeredness, so I just wanted to put that disclosure out there. And with that we will - do we have PQA in the room? Welcome.

So we'll move on to measure 0541, proportion of days covered. And I would ask that our measure developers give us a brief overview of the measure before we begin.

(Lynn Kazulo): Great, thank you so much. My name is (Lynn Kazulo) and I'm with the pharmacy quality alliance. We appreciate the opportunity to meet with the committee this morning. To my right is (Irene Ensia). We also have a few other members of our team that are dialed in. we have (Lisa Heinz), (Patrick

Campbell), (John Bedley), and (Ben Bannahan). And with that I will turn it over to (Lisa Heinz) to give a brief introduction of our measure.

(Lisa Heinz): Thank you. This is a sound check, can you hear me?

(Lynn Kazulo): A little bit louder, maybe, (Lisa), please?

(Lisa Heinz): Okay, thank you. Is this better? Is this better?

(Lynn Kazulo): I think we're checking to increase the volume here on our end.

(Lisa Heinz): Okay, how's that?

(Lynn Kazulo): Better.

(Lisa Heinz): Okay, great, thank you. We appreciate the opportunity to present this measure with three distinct rates to the committee. This is an important adherence measure used in the CMS Medicare part B star ratings program to evaluate Medicare Advantage and standalone prescription drug plans. According to the CMS 2018 impact assessment report, related to these three measure rates, health care costs avoided based on patient impact from 2011 to 2015 was estimated at between 4.2 and $26.9 billion.

This measure evaluates the percentage of individuals 18 years and older who met the PDC threshold of 80% during the measurement year. For diabetes medications, renal and angiotensive system antagonists, and statins, with higher rates indicating better performance. Using pharmacy and medical claims data that are readily available, the rates are relatively simple and feasible to calculate.

Recent studies support the growing body of evidence showing that adherence to medications for diabetes, hypertension, and hyperlipidemia is correlated with improved clinical outcomes and decreased healthcare costs. The evidence supporting the measure is directly applicable to the process of care being measured. The quality of the evidence is moderate. The quantity of the evidence is high, and the consistency of the evidence is high.

Furthermore evidence indicates that medication adherence for diabetes, hypertension, and hyperlipidemia remain suboptimal. Performance rates have improved since 2013, yet there is still a performance gap with opportunity for improvement and variations across plans for all three rates. Furthermore testing showed that there are disparities in performance for all three rates, based on age, race, LAS dual eligibility status, and disability status.

Based on feedback PQA received from health plans, and NQF recommendations to consider sociodemographic status risk adjustment, we convened the risk adjustment advisory panel. This expert panel selected patient factors for risk adjustment through a systematic review of the literature, and developed a valid risk adjustment model. PQA partnered with CMS to conduct a study using Medicare Part B data to test the final risk adjustment model.

For this meeting submission, the NQF scientific methods panel reviewed this as a complex measure and is satisfied with the measure's validity and reliability. This is a highly impactful measure that has been refined based on extensive user feedback and empirical evidence, using our consensus based process. In summary, this is an important measure that addresses a continued performance gap, and it's scientifically acceptable and feasible with strong feasibility end use. We thank you for your consideration.

Man: Thank you. With that I will pass it to our lead discussant, (Starlyn), to walk us through any additional comments and then move us into a discussion around the evidence.

(Starlyn): Okay, this is 0541, proportion of days covered, three rates by therapeutic category. They already reviewed a great description, so I won't burden you with that. I do want to make a point from a pharmacy standpoint, pharmacy claims doesn't necessarily give you the full vibrato of an EMR, so keep that in your mind. Also it's specifically around the Medicare prescription drug program, which is pharmacy claim fed. So there are some comments later on.

And we get - we're going forward. In five years that might all be different, so we'll see how that goes. This is four health plans and this is a process measure and they have been reviewed twice before, so they do - they did offer new evidence, and the new evidence mirrored a more modern approach of what we have been doing in healthcare and looking at diabetes and cardiovascular comorbidities and the fact that if you are not taking your medication that will have an impact on total cost of care and the well-being of the patient.

So I need to ask you if you want to go and dig back, because they have the old evidence, and they also provided the new evidence. I'm very - this is my primary focus right now for a quality assurance person for a public health and the - having an adherence rate will help us gauge what is happening in the communities on whether patients are receiving these medications, especially I the Medicare population.

And it could stem over to obviously other commercial programs. They had two new articles in 2016 that covered that association with ischemic stroke and non-adherence and Medicare Part - with type II diabetes, and then four in just 2019 that was added to the others. Does anybody want to…?

Man: (Jamie).

(Jamie): I reviewed this measure as well. I was curious, a couple of things. First of all I think it's interesting that you're talking about three diseases that in fact for those patients who have diabetes, all of these three actually apply to care of their diabetes. It's treating hypertension is very important in reducing diabetic complications and of course macrovascular complications are treated with statin. So it's a - it would be very interesting for them to be able to drill down more than just anti-hyperglycemic agents for diabetes. And I would wonder if they were doing that.

Woman: Thank you, that's a good question, so currently we are actually - it's more of a research study that we're looking into that interaction or intersection between those two (unintelligible).

(Jamie): And then the other question is why were individuals with insulin excluded from this macro - I mean, insulin is the major issue related to compliance, and so those patients who are on insulin are excluded from this analysis. Even if they're on insulin, it may be because it's hard to actually document how much insulin was actually taken, but still you can calculate the number of - the doses compared to - or you can calculate the total number of units that were prescribed in a given period of time and get some estimates.

The other thing is that those patients who were on insulin (unintelligible) are on insulin, a lot of them have other anti-diabetic agents at the same time, metformin and combination agents and various other things. So by excluding those patients who happen to be on insulin, you're also not measuring their adherence to the other medication.

Woman: Yes, and I think that's a good point, really. We currently have a measure that's in development that looks at insulin persistence, and to your question earlier because of the dates applied (unintelligible) with insulin adherence, you can't necessarily use the (unintelligible) methodology to get at insulin adherence. So we currently have a measure that just looks at insulin persistence, and that is currently underway.

Woman: So by pharmacy data, we go by the way of prescriptions are reimbursed to the pharmacies, and the NCPDP is the group that dictates that standard electronic claims process, and not - and the OMC is for medical claims, so while NCPDP has expanded those back in 1989 when we had over '90, we've expanded fields where we can kind of look at more than just a day's supply and refill too soon.

But we still are limited because in order for a payment to occur, some of those fields are not necessarily filled out. So you have to work within the mechanism of the claims data, past quality assurance director for Medicaid. So I'm very versed in that limited field. We've always wanted that inner operability between the EMR and the prescription claims, for adherence, so you can get a feedback for your information.

That's coming, but we need to be able to pull this now so that we can support the process of that three pillars of comorbidities and diabetes for that process. That's why they pulled those, in my opinion.

(Jamie): I understand, but in my MR, I have to delineate how many units are being prescribed for almost - for just about all of the insurances, they won't give you more than that. So they know whether or not the actual total amount of insulin units are actually being used or less, and that's still an issue that could be measured.

Woman: Well, going forward, the claims groups and the plans that were part of PQA's development of this, when they put it out for public, they requested for insulin specifically to be pulled because of all the different processes that are going in place, and all injectables, actually. And everything that's injectable right now in that claims processing world on the pharmacy side, that's all being brought to forth because of that very - some EMRs have that specificity.

Some EMRs do not. Some people are still writing paper prescriptions, and that's coming up in the reconciliation on the EMR side. When we fill it, we have to - day's supply is a standard of practice.

(Jamie): I see, but also I didn't see any exclusions for a non-insulin (unintelligible) injectables, which are being used more and more nowadays.

(Lisa Heinz): Yes, so can I comment on that? This is (Lisa Heinz). So if we can calculate regularly dosed injectables, we can calculate the day's supply, so that's reliable, whereas the insulin is frequently changed and then in terms of your question about why would patients on insulin be excluded from the measure if they're still on the other anti-diabetic medications.

And the rationale for that is because those who are on insulin, because of their frequent dosage adjustments, it may result in frequent holding of doses or in dosage adjustments of the other diabetes medications. So it's more of a - that's a clinical rationale. But it's a good question and it's something that we are trying to address with our insulin persistence measure.

Man: Okay, thank you. Ana, then (Don).

Ana: Thank you. I had several questions, and I'm still not completely clear on why insulin is excluded. I mean, I have type I diabetes. I've been on - I fit these criteria for many years, there was a lot of benefit with (Unintelligible) statins in prevention of kidney disease or slowing the progression of kidney disease, etcetera, so I mean, I don't technically fit the category of metabolic disease, but I don't know why that would be an exclusion. And then I have a couple of other questions too.

(Lisa Heinz): Because it cannot be reliably calculated. That's why we're developing a different methodology to assess that, to assess insulin.

Ana: Okay, so and then second, I have significant concerns about quality measures that require incentivize the use of statins, and I say this as somebody who's been on a statin, who knows the data very well, but there are a lot of concerns, a lot of significant problems that can arise from the use of statins. So you know, again, I feel like I've used this as an example many times in various discussions with this committee.

But my mother has this very significant adverse reaction to statins. She has - because somebody somewhere along the way, either a clinical guideline or EMR had this checkbox thing, answer that they were responding to, she has been prescribed statins of different types, in some cases by the same physician office, which was a different issue, but because people keep saying, oh, well, she's been on a statin. She meets these criteria. We need to get her on a statin.

And she has significant issues. She's not particularly astute in terms of understanding the generic name of a drug versus the class of a drug, etcetera, and so this is incredibly unclear. She knows she can't take a statin, but it's not like it's - the prescription says to the patient, this drug is a statin. So there's no

methodology for also incentivizing physicians to adequately and consistently screen for the adverse events of statins, which can be very nondescript.

So for incentivizing physicians to use statins like this, we also need to find a way to incentivize them to check very thoroughly and consistently for the adverse events associated with statins. And also there's a lot of evidence to show that statins aren't necessarily - they can be great and effective, and again, I've used a statin for about 20 years at this point at a low dose and I have great results.

But they do have an impact on mitochondrial repair, particularly for muscle tissue. There's been an increase in the incidence of type II diabetes for patients who have statins. So these are drugs that are relative to other drugs, maybe have a better adverse event profile than other classes. There's certainly lots of benefits. They're coming up with an incentive process that encourages physicians to sort of blanket prescribe those drugs.

I don't necessarily think it's helpful, and I think there are consequences that are very rarely considered, and particularly when you're dealing with the Medicare population. And there are lots of complexities that overwhelmed, elderly patients taking 15 different medications aren't always going to be able to stay on top of.

And then my final question is this feels really duplicative of the composite measure that we reviewed last cycle, and I'm just wondering what the rationale is there. I mean, it isn't completely duplicative, but there was also a requirement for statins as well as daily use aspirin and I can't remember, I think it was blood pressure. So anyway, I'm just wondering if we're coming up with different ways of requiring the same thing.

(Lisa Heinz): So if I could comment on that, this measure is not looking to start a statin, so it is assessing adherence to patients who are already on a statin, so it's not promoting new starts. And there will be patients who discontinue therapy due to a variety of reasons, and we don't anticipate that that would occur disproportionately across different plans for clinical reasons, and then we did include the exclusions of end stage renal disease, particularly because (unintelligible) are not beneficial in that patient population, and also frequent dosage adjustments.

Man: Yes, I think - you know, when I think about it, it's like when we looked at the composite last time, it was that - it was being done, and this is that thing that we said was so important to be done, that you continuously have that thing in hand throughout the year process. So that's kind of like how I think about the distinction. I'll leave the other comments either to the measure developer, or the committee if anyone has anything to say in response to the previous ones.

(Lisa Heinz): I also want to point out this is a prescription drug program plan, not the Minnesota composite that we looked at last session was for practitioners, and so this is on the other end in a retrospective way, and the other one was prospective in a sense, or current. So your - like Adam says, you're looking to see after someone has been put on the triple Aim, if they're staying on the medications.

And if the pharmacy part of the pillar is helping augment the physicians in making sure that the medications are being delivered and taken and not having problems, that's part of that assessment process.

Woman: So just to clarify, so I understand, you're measuring the electronic part of it, right? You're not measuring that the person, CME electronically get the

physicians to ping back the prescription to you, but you're not measuring whether the patient picked it up.

(Lisa Heinz): It's based on paid claims, administrative claims data, so there would not be a paid claim if the prescription was not picked up by the patient.

Man: Right. We had (Don) and then (Kevin).

(Don): Oh, there it is. It went on. It was flashing green, I thought that was good. Now it's flashing red, that's bad. So I'm sorry my (unintelligible) and this is extreme, I'm afraid. So this is measured at the part D drug benefit plan level, or at the health plan level like Medicare managed care, Medicare or managed care? Who are we targeting here?

(Lisa Heinz): So it is for the health plans, but it's actually for Medicare supplied at the contract level, so it's even more nuanced, so within the Medicare program, there are about 768 plan contracts, and that's for standalone PDPs as well as MAPD plans. And prescription.

(Don): So these are prescription drug plans, so like the one I have in my pocket. So tell me, I'm sorry, this is really stupid, what is the purpose of it for me as a provider or patient trying to have shared decision around medication optimization and adherence. What possible influence does my two-star plan, I couldn't find any plan with five stars.

Woman: The drugs, they're on the formulary and they're not too expensive.

(Don): But what is the - where's the - what is the use of it for improving the metric? Who's going to improve the metric so that more patients are adhering to their regimen in a shared decision with their provider?

((Crosstalk))

Man: Could you leave your microphone more than three…

Man: That means other people can't turn - there we go, can't turn theirs on, so when you're done speaking…

(Lisa Heinz): So the government wants to - is paying for your - part of your drug plan, right? You're paying some in, the government's paying some in. This is - okay, I'm putting on my old Medicaid hat. When someone else is paying the bill, they want to know that there's quality and it's going on. They want to make sure that - that's why we're here, the physicians and the health care providers and all of us delivering it. They also want to make sure that the medications they're paying for are being taken.

(Don): I get that. I mean, I think it's a tracking tool to see how we're doing as a country through our contracted services in Medicare and Medicaid. It's a great measure, but I don't see the actionable component that a drug benefit plan like mine has in - what their skin in the game is to help me take my medications. I can see how my Medicare Advantage plan or my care navigator or my provider can help me, but I have - the only contact I have with this company is when they charge me too much.

Man: (Katherine), and then (Lindsey), and then we have a comment from the telephone after you two.

(Katherine): So I'm going to put on my old Well Point hat. I used to run these quality operations for Well Point and was responsible for measure (unintelligible), and what the plan can do in this instance is they look and say, hey, I'm not

doing so great on this measure. Why is that? Oh, shoot. I don't have these drugs on my formulary. Oh, the drug's too expensive. My copay, the copay I'm charging is too high.

And then additionally the plan does work with this network and send out - the plan will then kind of cascade it down and look to see oh, these doctors or these groups don't seem to be adhering to the formulary and its measures, not a physician level measure, right? And I think there are a lot of reasons why there should not be a physician level measure, but what it inspires the plan to do is then to look to its network and say, maybe send them a note. Say hey, it looks like a…

Man: Let me just - because you're confusing me. So I have two cards. One of them is AARP. That is my Medicare supplement plan, right? So supplement one, that's AARP. It could be Well Point, it could be Humana, it could be a lot of them, right? It could be Medicare Advantage, whatever. This envisioned RX plus is who covers my drugs.

What I'm understanding you're saying is that this data is going to envision RX Plus, which AARP has no relationship to that I can discern. Well, what's the relationship? Do I go to - if I'm not on my drug, I can understand how AARP can call me, send me notifications, talk to my provider. But why aren't they the unit of analysis as opposed to this? This is a data-generating thing. This is the health plan that can actually influence my care.

Woman: Both health plans.

Man: Just a comment that every Medicare Advantage plan that works with our practice that we have attributed patients for sends me a routine report of which

patients are not meeting this performance metric, and they're held accountable as part of their star rating to the Medicare program.

Man: Okay, now we're getting somewhere.

(Don): Okay, so the accountable entity is actually AARP, not Envision RX Plus.

Woman: No, not for you. So you have Medicare supplements and - you have Medicare supps, plus a separate Part D plan, all right? So you're not in Medicare Advantage.

(Don): No, I'm not. I know that. I think it's too complicated for me to figure out.

((Crosstalk))

Woman: So it's the part D plan that is accountable, both part D plans and Med-Advantage plans are accountable in this measure, and in your instance that part D plan is going to be measured. They're going to look through this, the evaluator, the government's going to look and say, hey you're a part D plan. How come the beneficiaries that are in this plan don't seem to stay on their drugs?

Man: Because Humana's no damn good. I mean, if part D doesn't call me up, or my provider…

Woman: Well, it might. Part D might.

Man: Somebody ought to draw up the logic says, because as somebody who spent - literally had to hire a consultant to figure out what to do, this measure makes zero sense to me as you're discussing it. That said, I'll let it pass, but

somebody ought to draw it out, because I'll tell you, the confusion out there for people like me to have to hire a consultant to figure out what's better for me and what's - and I looked at the part C and nobody had good rating.

The ratings were all two stars, three, whatever. There were no amazing plans in Massachusetts. So, and I was told by my consultant that that's the way it is. So that knocks off the chip. The other thing I want to ask, is somebody going to talk about the risk adjustment at all anymore, or we just accept it?

Man: Yes, it's going to be later. We're not there yet.

Man: Just wanted to be sure. Thank you. So fine. I just don't get it, but I'll vote in it.

Man: Okay, (Lindsey) and then William Taylor.

(Lindsey Bachford): So I do have an evidence question, and it relates to how - so there has been published evidence that every other day statin use might be just as effective as daily statin use. Is the measure looking at whether the medicine is taken as prescribed or whether it's taken daily?

Man: It's where they got it, right? It's a proxy for adherence.

(Lindsey Bachford): Yes, so it is - it does look at how you're taking it every other day, and it's prescribed daily you will pick it up less often.

Man: So I'll just say, and I can only speak for the disease state that I come from in HIV, we continually have patients who routinely pick up their medication and don't take it, because they don't want to have someone to have a conversation with them that they're not picking up their meds. We constantly find folks when we switch meds who will bring us bags full of unopened prescription.

So while I know that doesn't sort of make this sound really great, because it's kind of showing it doesn't quite get you an adherence, it's more that was this routinely being picked up over the time period we expected the individual to be on it? But it does kind of get at that. I mean, I like to think of in the sense of the thing about pre-exposure prophylaxis where people don't take it as prescribed and do spread it over time.

This would show us that individuals may potentially be using that medication that way, so it can show you that, but I don't - and correct me if I'm wrong, it's not showing that you took it every day.

(Lindsey Bachford): So it is - the measure is calculated based on the way it was prescribed. So if the medication is prescribed to be taken daily, the calculation uses the day's supply field, which is determined by the quantity dispensed and the instructions for taking it. So if the patient is intended to take it daily, then it is assessed to look at are they taking it daily. If it's prescribed to take the medication every other day, then it would be assessed are they taking it every other day.

Man: So real quick we've got to go to William Taylor on the phone. He's had his hand raised for quite some time.

William Taylor: Thank you. I'm glad you're going to get to risk adjustment in a moment. I have something to say to them as well on disparities, but just on the technical question about measuring adherence and its association with outcome, I mean obviously adherence is critically important. There's no point in prescribing these medications or having studies that show that they work if patients don't take them.

So we all understand the importance of adherence, but measuring it for all the reasons that have been identified and another one I'm about to mention is treacherous. I know, what I'm about to mention is that it's been known for more than 30 years from a control group in the coronary drug project in the 1980s that people who take their placebo do better than the people who don't take their placebo.

So medication adherence identifies different people from the people who are non-adherent in the so-called unmeasured confounding ways that are hard to get a hold of. So we look at adherence and show the studies that show people do better when they adhere, of course they do. That's how these medications work, but figuring out what that relationship is, is quite complicated, and I'm going to say more about that when we get to risk adjustment.

Man: Yes, I think a question that I have too kind of stemming from that, so I know my medication is delivered to me every month even when I don't ask for it. It just comes to me now, and because it's chronic medication, I'm going to be on it forever, I called and said, hey, just keep sending it, right? So that I don't have to have this conversation. So in theory then I would show up if that medication were one of the three we're looking at, then I would meet this measure even though I'm not actively going to get it.

Man: Correct.

Man: Okay.

Man: (Katherine), and then (Jamie).

(Katherine): So question for the developers, how do you account for - I don't think that you can account for this, I mean just kind of thinking out loud about how this

applies to the measure. We know that certainly for statins there's a fair number of patients who need to stop taking the statins. So they develop statin myopathy, aches and pains with their muscles.

And so those patients are probably a smaller percentage, but do you know what percentage of patients who take statins have this problem? Is it enough of a problem that it's going to be a problem to measure? I mean, I think that'd be less of a problem with the diabetes meds. Some diabetics, we're going to be able to say type II diabetics, they can come off their meds and be diet-controlled. That's probably a pretty small percentage. But you know, that - I think this would be a bigger issue as a physician level measure.

Man: … stop describing indications.

(Katherine): Right, but those would be failures in this measure, and I'm just wondering how much of a problem it is.

((Crosstalk))

Woman: Go ahead. Go ahead, (Lisa).

(Lisa Heinz): Oh, so the anticipated rate is not 100%. However, and so it is expected that patients, that some patients would discontinue therapy for a number of reasons over time related to myalgia and things like that, there's always the therapeutic alternative across the class. At a population level again we wouldn't anticipate these discontinuations to adversely impact one plan over another.

Man: (Jamie)?

(Jamie): Yes, I have a question. It's clear that I understand how this measure would be used for quality improvement in a whole variety of different ways, but how is it used for accountability? And it seems to me there are a lot of treacherous aspects to this.

Man: So we will hold that question until we get to use and usability.

(Jamie): Okay.

Man: Yes, thank you though, and (Faith), and then we'll go over to (Victoria).

(Faith): Just a general question thinking about the evidence and things that have been looked at, because in the Medicare population, it seems like this is very good on the way and it's showing us in the correlation in the numbers that it is doing the right thing, but one thing to just consider, or have you considered when you start looking at other lines of business or other populations the fact that they receive a lot of their medications for free like in clinics.

How are we going to think about those populations when this measure is - you know, the population is expanded? Because one, and this recently came up where a clinic is like, well we have excellent A1Cs, we're talking about diabetic medications, but A1C scores, but the compliance or that 80% adherence, they do very bad at that measure because they get their medications for free or from somewhere else so they give them samples or what have you. So as it does expand, is that something that has been looked into? Just thinking about how does that range?

(Lisa Heinz): Yes, we wouldn't have this ability unless they were incorporated into the claims data, so that is a point to consider, and you'd be able to capture that perhaps with an all payor claims database in those populations, and this - these

measures are used in some Medicare or Medicaid programs. There is a need to improve adherence in other patient populations as well.

Woman: And to add to (Lisa)'s point, so there were some studies that didn't just look at Medicare, there were studies that actually looked at commercial, and the results, the trends were similar. Sometimes the magnitude is slightly different, but the trends were similar in the impact of adherence to some of these outcomes.

Man: So I'm just looking at our clock, so I'm going to ask that we stay razor focused on evidence so we can get through this measure before people have to get on planes. So I believe it was (Victoria) and then (Kim).

(Victoria): Okay, so I'll be super quick. Is there no way to put in an exclusion, or is the problem that you - an exclusion I guess to this issue of someone has adverse event, thus can't stay on medication, there's no way to put that into your measure development?

(Lisa Heinz): They're measuring a continuum. If they started on it and if they stopped, if they started and they took a month and they started having myopathy, then it's discontinued. At the end of the year there is 11 months of no data, so they would not be measured.

(Victoria): They couldn't be - I guess, so 11 months of no data, they are lost to follow up? They couldn't be that outcome happened at month one, and that's their…?

((Crosstalk))

Woman: For PDP plans all they have is pharmacy claims, so they won't know if there is an adverse event or - a Medicare Advantage plan might know that, but we won't know that.

(Victoria): So I guess that same would happen if the patient died, right, then?

Woman: Correct.

((Crosstalk))

Woman: Well, not if they died, but…

(Lisa Heinz): So the measure is calculated retrospectively, so if they died, that would be captured. As far as exclusions, some of the challenges are the data that the plans have access to. The way that the measures currently are used, CMS in that retrospective nature can apply those exclusions when they are calculating the rate. So we have the ESRD exclusion and others.

(Victoria): And there are no adverse effect, there's no unintended consequences getting extra prescriptions that are never being used?

(Lisa Heinz): We have not identified that to be an issue.

Man: Yes, so as it's written there's no exclusion for adverse effects for either the statins or for the RASA medications. There are ICD10 codes that a provider could put in when there's an issue, and that could trigger an exclusion, but yes. It - yes.

(Lisa Heinz): If we got an ICD10 code, we would love to have that in the pharmacy because then we could pair data together and we could help this process. We're still in

those discussions because everybody wants to hold onto their own data and not share that, and not give us those ICD10 codes, but someday that's going to happen.

(Victoria): And what happens if the prescriber denies it? So the pharmacy pings the prescriber and says we want a renewal on this medication, if the prescriber then says deny these two, adverse event, will (unintelligible) take it not been to clinic or what is the…?

Woman: So again the assessment is - the unit of measurement is at the health plan level, so again we do not expect the measure rate to get to 100%, and because it is a population based level, at the health plan level, those instances where there are appropriate discontinuations as (Lisa) had mentioned earlier, we wouldn't expect that to impact any one plan disproportionately in comparison to another.

So that does happen and prescribers are making the appropriate decisions based on their clinical decisions. So if a patient needs to discontinue a medication because of an adverse event, we would anticipate that that discontinuation would occur. But again, at the health plan level we don't think that that would be - would affect one plan's rate disproportionately compared to others.

Man: (Kim)?

(Kim): One of the things that was referenced in the evidence was really the conceptual relationship between this more of a process sort of measure and outcomes measures, and I just want to emphasize again that we really need to have it as close as we can to outcomes measures, because ultimately that's what really matters and what we should be focusing on.

Man: Any other comments on evidence? Great, let's move to the vote.

Woman: So we're voting on measure 0541, proportion of dates covered, rate of therapeutic category, and we're voting on the evidence. The options here are moderate, low, and insufficient. Waiting for two more votes. Give me just a second here.

(Starlyn): If you can open the calendar invitation it should be in there. Okay, so 14 votes for moderate and six voted for low, so we're just converting it over to percentages. It should be 67, 66, 67%. So where it passes.

Man: Any comments on performance gap? Okay, we can take a vote on that. Wait one sec. we have a comment on the phone. (Bill Taylor)?

William Taylor: Yes, thank you, I was still on evidence trying to get recognized. I realize it's difficult with that many sitting out here. when we think about the evidence, when we have a continuous measure like adherence, and we draw an arbitrary line at someone like 80% and say on one side of 80%, that's okay and on the other side it isn't, it's pretty important where that line is, because we get very difficult results if we drew that line at 90% or 70 or 60.

Is there evidence to support drawing the line at 80 and that you know, people at 79% are somehow inappropriately in this is a deficient category, and people at 81 are appropriately into that fine category?

(Lisa Heinz): Thank you. I think that's a good question and one that we're currently actually exploring from a research perspective, but when we first looked at this measure systematic review of previous measures showed that PDC greater than I recall to 80 was used in over 90% of studies between 1980 and 2004,

and so we just went with what the current literature review said and based that measure off of that. But as we go along with the measure, we are currently looking at the different thresholds, and if there's a better threshold, and that's something I think from a research perspective we're currently exploring.

William Taylor: Great, thanks.

Man: Any comments on performance gap? Okay, we can move to the vote. It doesn't have to be.

Woman: Measure 0541 on performance gap, the options are high, moderate, low, and insufficient. (Unintelligible), just waiting on one more vote. I'll close it. So for performance gap, three committee members voted high, and 16 members voted moderate. This measure 0541 passes on this criteria.

Man: Thank you. Now we move on to scientific acceptability. This was reviewed by the scientific methods panel. Did you have anything else that you would like to add, (Starlyn)?

(Starlyn): I just wanted to point out that the scientific acceptance panel had - was - had a question about the Medicare population, and so they - in 2017 they resubmitted 100% of the CMS Medicare claims, which offered a better reliability and validity both because it was all consistent data that didn't - that was better than the 5% they had previously submitted into the process. So they - the committee voted for reliability, one high, three moderate, and one low. So overall it was moderate.

Man: Any additional questions or comments related to reliability or validity?

(Starlyn): I will remind you this is prescription claims data.

Man: So when thinking about reliability, just to be clear, we're going to - the vote that we're going to take has to do with accepting the scientific method panel's assessment and recommendation here, versus us having our own discussion.

(Starlyn): So just a quick process point, if you want to accept their recommendation you vote yes, if you don't, you vote no. the majority of the committee votes no, we would then have you discuss and vote to make your own recommendation.

William Taylor: Just for clarity, at what point does the recommendation around the risk adjustment get discussed?

(Starlyn): Validity. Next, on validity. Yes.

William Taylor: Not here.

(Starlyn): Not here.

William Taylor: Okay.

Man: Get ready.

William Taylor: Oh, I like risk adjustment, don't get me wrong.

Man: I think we can go ahead and move to the vote.

Woman: We have the option to vote yes or no, do you accept the scientific methods panel's moderate rating for reliability? So we had 19 committee members vote yes to accept the scientific methods panel's moderate rating for reliability. And we're going to vote on validity next.

Man: So we'll discuss, right? Just making sure we're doing the process right. Any comments or discussion around validity?

(Starlyn): The committee gave this a moderate vote, and they based it on the base validity was kept at a two ways construct, and empirically there may be some questions about the differences between the diabetes medications and then the RASA and there were some concerns about leaving out the combination therapy, but I will let you know that that combination therapy is not prescribed for hypertension. And that the RASAs were used as a hypertension indicator, not a chronic heart failure.

And then there were some issues on the statins again, on the base validity. But this - there were ors. This is an and. You're looking at a population, and we're looking at people that take the medication for diabetes, oral medication for diabetes, people who are taking a RASA for hypertension, and people who are also taking a statin in a year's worth of time. You have to think about that. Right?

William Taylor: I thought we weren't supposed to consider base validity for an endorsement of a maintenance, of a measure already approved.

Man: That is correct. So just for a point of clarification, the correlations that were demonstrated here were using (unintelligible) correlation and coefficients, looking at a core level, so those would not be considered a base validity assessment.

(Starlyn): So I think at this point maybe PQA could talk about how they look at base validity through your PQA committees, because there was confusion on the process of how you review this internally, or is it at another time? (Dan)?

(Dan): That piece of evidence doesn't necessarily need to be considered by the committee.

(Starlyn): Okay.

(Dan): You may choose to discuss it if you wish, but it's not necessary as part of the maintenance of an endorsement.

Woman: Okay, I have some comments though from validity itself.

Man: Any other comments or questions related to validity? Now would be the time. Risk adjustment, (Don)?

(Don): I need to put my card up. No, I'm hoping that we can be satisfied that the risk adjustment takes adequate account of the issues that affect safety net organizations in America and include issues of social determinants, because if they don't then it's not for this purpose real risk adjustment, so a little detail about that would be great.

And beyond that, I say this every time this comes up, whether it be readmissions or whatever measure NQF is recently endorsed, if the measure is not accompanied by a requirement that the strata of performance be included, then it nearly obfuscates the issues of inequity in America.

So I'm all for risk adjustment, the payment and penalty and all that, but it's got to be accompanied by clear demonstration that the data have been stratified, and that black moms are dying more than white moms and that Native Americans are not getting the drug adherence that white Americans are

getting. So without that I won't (unintelligible) measure this risk adjustment (unintelligible).

Woman: So to the second point about the stratification, the measure itself does require that the variables that are used to stratify as part of the measure (unintelligible).

(Don): That's clarified and reported?

Woman: By the implementer, yes. By whoever's interested at CMS. So we as part of our recommendations for risk adjustment, we do state that the implementer makes this information available to health plans, and should be stratified by the variables that we're using for risk adjustment.

(Don): Okay, that's major, and if that's what's happening, that's probably the first time I've seen that. So that's really great.

Woman: So that's our recommendation. How the measure implementer actually uses it, so the specs are set up to require that they be stratified, and then CMS has asked the implementer to explain how they are going to (unintelligible).

(Don): Okay, thank you.

Man: Any other questions or comments related to validity?

William Taylor: Yes, it's William Taylor. Can I say something on the risk adjustment?

Man: Yes, sir.

William Taylor: There's a danger once you start dinging anyone, plans, clinicians, whomever, on the issue of adherence, that one approach is to make your services less hospitable to the people who are more challenged to adhere, right? And whether that be people who have low income, people with low education, people with mental illness, people who don't speak English, there's a grave danger of this. And even mathematically a valid risk adjustment won't necessarily take care of it, so it's another reason why this is such a really challenging, or even treacherous area to get into.

Man: Thank you. Final question or comment? (Bill)?

(Bill): So when you look at their validity testing, they did convergent validity in all three of these, and they paired them with other process measures, and the coefficients were around 25 for the oral diabetic medications, the RASAs, and it was a .35% medications, I'm not sure that's - I mean, it's not stellar in my mind. I mean, (unintelligible) statistically significant net correlation but (unintelligible).

Woman: So I think part of that is to consider that this is measuring a relationship, not just in construct, but adherence wouldn't necessarily be the only reason why someone's blood pressure (unintelligible), that the plan that is performing really well on the blood pressure may also perform very well on adherence.

So it's that the relationship is in the right direction and it is statistically significant by conventional standards. .3 is moderate, .5 is high, for correlation, for (unintelligible) correlation. But again it's because there's too many other factors that we go into the low blood sugar or that kind of the measure including lifestyle choices, things that adherence wouldn't necessarily capture.

Man: All right, so I think we can move to our vote, and again this is a vote whether or not to accept the recommendations of the scientific methods panel, yes meaning you do accept it, no meaning you do not accept it and would like for the committee to (unintelligible).

(Starlyn): Okay, so again you are voting on if you are accepting the scientific methods panel rating for validity of moderate, options are yes or no. Okay, and I believe we have all votes. 18 people said yes, and two people said no, so it moves forward, if you go forward on the next criteria.

Man: Move to feasibility. (Starlyn), if you'd like to - can you turn on your microphone?

(Starlyn): Their comments were there was no concern for any of the feasibility because this is through a prescription drug program plan's data, and it's been - actually been in place for 10 years for those of us who are in the pharmacy world. We've looked at the proportion of days covered in not just these three areas but in other area of medication use.

I will also say that this helps identify that area of patient care that pharmacists need to extend in order to have a feedback mechanism with the physician to let the physician know what's going on, so that's one of the triggers that you use to say under Medicare Part D, you have medication therapy management programs, and the pharmacist can work with the physician and the patient together to work on that adherence. So it's been working for a while. I know it was considered moderate, the feasibility.

Man: Yes.

(Bill): So I wanted to raise an issue that we've had with this metric, because again I think I have four different Medicare Advantage plans and their patients and they're all sending us reports and they all have clinic payment programs. They're getting star rated, and so they're holding us accountable for the adherence, and when we looked at the chart (unintelligible) the issue of dates where on our campus there's a big VA hospital.

We have medium health service. We have lots of other places the patients could choose to get their medications, even when we prescribe them. They choose not to use their plan, and it's been very, very frustrating for us. So we look in the chart, we can see the medication is ordered. They picked it up on the $4 Wal-Mart plan when they're doing their grocery shopping, so - or they see their PCP at the VA twice a year and get all the prescriptions filled at the VA, even though they're seeing our doctors through the department of planned (unintelligible) health services (unintelligible).

So there are some challenges around the measure that we have found frustrating. It frustrates our admissions when I call them and say hey, Humana says your patient's not adherent, and I wrote a prescription for them every month and they're taking it (unintelligible), but they tell me they're getting it at Wal-Mart, and I mean, we find evidence that they're taking their medicines. They just aren't in the claims as a claim because it's very plan specific. So it's one of the limitations we've noted.

(Starlyn): I agree with you. I've seen the very same thing, having especially in a Medicaid program when people come in and out, and I'm - right now I have the city of Chicago and we have a cardiovascular diabetes prevention and a maintenance program going on, and many of the employees of the city of Chicago have multiple - they get some meds from VA or they get tired of maybe of Caremark and go to their favorite pharmacy and pay $4 for the

metformin. So right, if we don't have the claims for it we can't follow it, but this is still looking at - so when the healthcare plan is trying to contract with the PDC, what the PDC's performance is in that process. So I hear you and…

Man: Any other questions or comments related to feasibility? Okay, we can move to the vote.

Woman: So (unintelligible) feasibility is now open, measure (unintelligible) replies for one sufficient of the others, rate by therapeutic category. And so feasibility, the options are high, moderate, low, and insufficient. Just waiting on one more vote. All 20 committee members have voted, two voted high, 16 voted moderate, and three committee members voted low for feasibility. The measure passes on this criteria.

Man: Thank you. Lastly use and usability.

Woman: There are a lot of comments here. Again some people were still very concerned about the exclusion of hospice and the renal disease people but they wanted that to be incorporated in the measure and that to them impact of the usability. Based on my research the feedback from the users of your measure elected to have that - have those, the Hospice group and the ES end-stage renal disease patients because they're fluctuating that these three medications and whether the in that end of life are you actually part of that - be part - should be part of that measurement for adherence in the hospice group. So I just wanted to not ignore that person's good input and but most - it's a very usable measure based on claims data. And CMS has ten years of experience with it.

Man: (Don)?

(Don): Hi. And could you explain to me again how it's usable by Part D drug plans to improve quality of care? I just looked at the Massachusetts thing and it looks like there's only one five-star rated plan in Massachusetts and the main reason for that seems to be the weighting of these particular type of drug adherence measures in the overall score. So tell me how I as a patient know that the Part D plan I'm contracted with is improving, can include my care as (unintelligible)?

Woman: There's another - there's a number of interventions that can be used to improve adherence including medication therapy management programs.

(Don): But that's not the question. I want to know how my - how envision health, envision RX, the Part D plan what it does to improve that measure, it specifically? That's the usability issue. I know I spent my entire life trying to improve medication adherence and use so I don't need a talk on that. I want to know how Part D plan specifically are using this measure to improve care for people.

Woman: I think she was…

(Lisa): Perhaps I don't understand the question and I can defer to my colleagues. I don't know about that specific plan.

(Don): Any plan, any Part D plan, tell me what a - any specific Part D plan how it uses that information to work with providers and patients to improve med adherence?

Woman: (Lisa) if you can continue with that I - she is getting to that. She's giving some examples of the interventions that the plan put in place to be able to improve on the measure. Yes (Lisa)?

((Crosstalk))

(Lisa): Okay thank you. There is a yes so there is outreach to patients through medication therapy management program. There can be reminder devices that are provided. Lower co-pays is another example, Even financial incentives can be a flexible way in some innovative models in Medicare Part D plans. There can be collaboration throughout pharmacy networks or provider networks with performance reports.

And this is known and in the literature review we have some citations supporting that describe the interventions that can be used to move these measures. And in fact performance has increased over time so in addition to there still being a performance gap I think one of the biggest things is that over time the measure rates have improved and Medicare Part D in their impact report this is one of - these are these three rates are some of the most impactful measures in terms of at a patient level looking at impact and then cost to the overall the Medicare for Part D program.

Man: (Lynn)?

(Lynn): I just want to make a quick comment since I mean I applaud (Don)'s concern about how does this help patients. But speaking of somebody who takes at this point 16 different meds one of the biggest - I'm like religiously compliant and the only reason I'm not compliant is the complexity of getting access to the medication.

So if a pharmacy benefit program or insurance company is incentivized to make that as seamless as possible and as easy as possible that's a good thing. I have issues with the statins part of this measure but in terms of like the

potential benefit for patients if that is indeed what happens because of this that's a very good thing.

(Lisa): Yes access to care…

(Don): I'm going to take this on faith and not pursue it any longer but if there are citations in the document that I missed that provide evidence that Part D plans specifically those actions have had an impact on the performance measure I'd love to see them as opposed to all the other things, the providers and health plans are doing. But I'll let it pass for now. That I've just never seen that evidence or encountered it at that level and my in my own family never received a call from a Part D plan.

(Lisa): That's unfortunate. Yes I'd be happy to follow-up and I can also look it up right now.

Man: (Lindsey)?

(Lindsey Bachford): So I think overall the benefits still outweigh the risks but I disagree with the preliminary reading of high due to I think potential unintended consequences that Dale was hinting at earlier. While it is a health plan level measure I think especially in the MA world the health plans are taking it to the physician level and penalizing physicians as a result. And that is the piece that I do think is an unintended consequences measures like this. Patients could have good adherence but since it's using claims level data getting medication to the VA, using samples or using a $4 list are not accounted for in this.

With one of our particular MA plans our existence on the plans is contingent on our star level rating. And based on how they do the star level rating in a given year medication adherence could be quite important. So I think it has

unintended consequences that aren't accounted for here. I do not think it is high. In-between moderate and low and on the waiting of how important that is I see the arguments that yes it encourages low co-pays and things like that but the message back to the physician is if I want to view - score it appropriate by my MA plans I need to tell the patient to fill it out - the pharmacy their insurance wants them to get it from, not what might be best for the patient. And I have a problem with that.

(Don): Yes and the question I had too sort of relates to how pharmacists are reimbursed as well because I - does this correlate down to that because I know the pharmacy that's co-located with our clinic like if I had a dollar for every time this man screamed with his reimbursements of how he got dinged because he had a low denominator of patients where one didn't pick up the meds which dropped him beneath 80% and it cost him $20,000.

So it's like that's where I was sort of looking at this trying to figure out is this driving that? Is that how it's coming down from the health plan to affect adherence which is causing our pharmacist to basically like lose money and to then inspire adherence. But if the denominator is that low and you only have one person it seems kind of like pretty strict to kind of have that kind of punishment and penalization on what they're being paid for one person not picking up their medication.

Woman: So that's a valid point and I'm glad you brought it up and I didn't because I'm usually in trouble for, you know, being the capstone for the abuse of individual pharmacists, not necessarily - and independent pharmacies are being raked all over the coals because they are not part of a narrow network of some of the other big boys out there that are using our quality measures and process measures against us.

I know as a member of PQA that I brought this to the - I'm the one in the audience the raises the hand when the PBM, Pharmacy Benefit Manager people are on stage and asking that very question. I - this measure in particular is a drill down in order to identify those patients and identify the plans that are doing well in that process and hopefully guiding the - it's health plan is the view saying oh, well these people aren't doing this very well. We're not going to re-contract with them.

However the on - the very unintended consequences is just coming back and being used in the DIR claw back (feed) and that's very - that is happening. It's happening on I don't know through - it depends on the plan and who's the PBM that's part of that plan and what network contract they have with those pharmacies.

But the ability to have the gag clause removed from individual pharmacists we can now talk to our patients and say, "Hey did you know if you just pay cash for this it's $4?" And we didn't have that a year and a half ago. So it is - it has the potential of unintended consequence with that but I don't know how PQA or any other quality measures developer people, any of us can go back to the user and ping them.

Some - to some extent that's going to happen in the long run because they're now pharmacy benefit managers are not going to be treated like they're all alone. They're going to be pulled into the insurance world and each state by state they're going to be held accountable and be more transparent. So they're - they need to be part of solving the problem and I think that's coming to fruition. Three years ago that was a different issue but now, very well - I'm glad you brought it up so thank you.

Man: Measure development then we'll go to (Kathy).

Woman: Yes so we do provide guidance around the use of these measures - of this measure and indicate that it should be used only when you have a minimum denominator size of 30 or greater but appreciate the feedback from the committee.

Woman: That partially answered my question. But I think, you know, the whole discussion we're having great now is really just calls out the problems we have with our terribly fragmented healthcare system. And, you know, this measure would not be appropriate at the physician level. But, you know, as discussed, you know, probably appropriate at pharmacy level, at the pharmacy plan level but there are some problems with it. And I just wonder if NQF, you know, would consider, you know, putting together some sort of a kind of a statement, you know, that says look, you know, we have measures that are plan level measures, and we have measures that are, you know, at different levels. And you really can't extrapolate them as the plans do to their network positions. You know, they - I think it's fine for them to yes reach out to the network physicians and say, "Hey, you know, we see this patient, you know, give you a list. These patient you're not to be adherent, take a look at it," but I don't know that they should be using it in terms of your compensation and they shouldn't be.

Man: All right well thank you. So the response to that has been the way that our measures specifications are structured (unintelligible) if this is specified at the plan level and endorsed at that level then that's where it's supposed to (unintelligible).

Woman: Right but that's not what I was saying. But I - that's clear. What I was saying was it might be useful if NQF were to make a statement on it that you should not use it as a (unintelligible).

Woman: So this comes up quite a bit because we endorse measures for the level for which they're specified but when they go out there in the world they can be picked up and implemented for other levels. And so one of the things we're trying to do through our external database, the quality positioning system is be very explicit on where it's at. This is the level for which was endorsed. It's almost a statement that may come out. Other recommendations from other committees is that they explicitly say that in the report for all the measures that goes through. So that might be something that's coming perhaps in the fall.

Woman: Okay sorry. How much of that thing, you know, I understand that like this issue of what the patient does or doesn't do or where they fill it so health plan has no more control over that the - of that than the physician. So it's kind of the same problem. It's just a problem no matter who - because you (unintelligible) I mean the health plans can't say just like you can't tell them to go - not go to Walmart we can't tell them to not go to the Walmart either.

So it's still kind of the same issue no matter what lens you're looking at it from. So and to that same end unfortunately speaking like working with a lot of the NQF endorsed measures there may be measures but if you're a health plan trying to help drive outcome for your members the measures in which you are going to have to use are typically process and if you're really trying to push the envelope, you know, like really drive that and help to get better outcome you kind of have to not used standardized - you know what I'm saying? So it's kind of like we're - it's the collaboration of all the pieces involved like how do you get to using standardized measures that are truly driving the outcomes? I'm - just a side note but I think the issue is for whoever you're looking at it for.

Woman: So this measure is a list of measuring is the computer system working right? Is it pinging the right people and is the script getting refilled and then delivered to the patient? It's not actually - it's probably not measuring what would truly affect outcomes.

Man: Just one last question. So I am assuming from everything that's been said in favor of this measure which I think is a okay tracking measure for how we're doing in this country, I'm assuming that the folks who work in Medicare and Medicaid have worked with and have - are currently equally involved in a coordinated approach rather than what seems to be I'm going to get a call from Envision RX, I'm going to get a call from AARP, I'm going to get a call from the patient portal at Mass General Hospital and I'm going to get a nag from my wife to take the medication.

So that's not good. I mean, when I get a phone call from Envision Rx I'm assuming it's going to be about money and payment and I don't answer the phone. So of all the things that would be a powerful lever this is probably the weakest if that's the mechanism.

And so I'm hoping and praying that there is a coordinated effort and I would just hate for us to be seen as enabling a fragmented effort by improving a measure that's one tiny - I would say the attributable fraction for this to a Part D plan on getting providers and patients to work together and share decisions and self-management is on the order of 2%. I'm just guessing but that's my estimate.

Woman: So 3% to 5% in diabetes right now, 3% to 5% in diabetes right now just depending on what part of the country you live in, in shared decision-making. We're under - we're not meeting the needs of our - of the population that's out there. But there is more coordinated efforts. There is a huge - CMS and CDC

are pouring a lot of money into team approach to where were involving everybody. And so when the pharmacist is getting dinged that that then becomes part of that communication with the patient, "Hey what's going on with the patient?" So there is a big effort on that. It just hasn't got there yet.

Man: Our pharmacy closed and moved as a result of it to expand the denominator of patients is what happened. So I hear you on the sort of unintended consequences of how it goes down because now a specific chronic position no longer has its pharmacy in-house to pick up meds. They had to move to expand the population because they were getting so many penalties for their reimbursements as a result of not hitting their mark. (Unintelligible) go ahead.

Woman: Yes so just a couple of comments. So I agree I believe that, you know, the importance of the coordination throughout the system certainly was important and I think that's being heard at multiple levels and efforts are being made to improve in those areas importantly. And then also on, you know, regarding the way that measures, kind of that triple down affect with the physicians pharmacy PQA currently is we have pharmacy level measures in development.

So, you know, we do appreciate the feedback from the committee and, you know, during about the use of our measures through various feedback loops. But, you know, this is one way that we are looking at opportunities to have measures that are developed for the intended use of accepting pharmacy level performance.

Man: Any other comments related to use and usability? Okay I think we can go ahead and move to the vote.

Woman: Poll is now active for the use criteria on measure 0541 for portion of days covered rates by therapeutic category. The options are passed and no pass. Waiting on three more votes. After 20 committee members who voted on the use category 14 selected pass and six selected no pass. That's 70% and so the measure passes on this criteria.

The next is usability. The options are high, moderate and low and insufficient. And this is a measure 0541. We're waiting for two more votes. One more vote. Using 14 as quorum we have 19 committee members who have voted so far. Three voted high, nine voted moderate and seven voted low on the usability criteria. So we have 63% of committee members voted high and moderate and 37% voted low.

Woman: It passes.

Man: Thank you. Related and competing measures?

Woman: So there are two other adherents medication measures 1879 and 1880. One is antipsychotic meds and the other one is mood stabilizers. But those are for different medications and different disease states. There was one concern about duplicating efforts from a physician or insurer base measure. It was shared in the above usability but it applies more in competing measures. There was a concern that just at the pharmacy level happening and then the physicians being a similar but I think we've hatched that one out quite a bit already.

Man: So I think we can move to overall sustainability or suitability for endorsement?

Woman: The voting is not active for overall sustainability for endorsement. The options are yes and no. For overall suitability 60 committee members voted yes and four committee members voted no. That's for measure 05414 for personal days covered bridge by therapeutic category.

Man: So now we are at lunch.

Woman: Public comment and then lunch.

Man: Oh public comments.

Woman: The lines are now open for public comment. If you have a comment on the phone please speak up now or submit via chat. Okay no comment so we will now have lunch.

Woman: (Unintelligible).

Woman: Okay 1:45 please be back. Thank you. Okay everyone we're going to get started in a minute or two. If the developers for the next measure want to come at the table feel free to do so.

Man: So welcome back from our very long and extended lunch. Thank you everyone. What we'd like to do now is move on to our next measure, Measure 30591 time screening for Hepatitis C virus. We have measures stewards PCPI here with us. If you guys would like to give us a brief overview of the measure and then we'll move to our lead discussion.

(Jamie): Sure, thanks. My name is (Jamie). This is (JR). We have (Kerry) on the line and also Dr. Wong and I'm going to passes over to (Kerry) to give a brief intro.

(Kerry Fy): Hi. Thanks (Jamie). This is (Kerry). Can you all hear me?

Man: Yes.

(Kerry Fy): Great, just wanted to make sure. Okay so thank you so much for hearing us out this afternoon. Like (Jamie) said, my name is (Kerry Fry). I'm a Program Manager in measure development operations at the PCPI. On behalf of our team I am pleased to present the first of two ECQM measures for your review today. The one time screening for Hepatitis C virus in patients at risk measure evaluates the percentage of patients aged 18 years or older with one or more of the following. They have a history of injection drug use, receipt of blood transfusion prior to the years 1992, receiving maintenance hemodialysis or a birthdate in the years of 1945 to 1965 received a one-time screening for Hepatitis C and (infect) a virus infection. This measure was approved for trial use in 2016 and we're bringing it back to you today for your consideration for full endorsement.

As noted in our submission of the 3-1/2 million people living in the US with Hepatitis C virus only about half of them know their actual infection status. While (HCVL) causes acute infection most people are asymptomatic. Studies have shown that the treatment of HCV infection is cost-effective and that 90% of cases can be cured if identified. If left untreated infection can become - can become chronic putting those infected at higher risk for liver disease such as cirrhosis, (hypathothelial) carcinoma and extrahepatic complications. Despite the success of treatment literature shows that screening rates for HCV remain low. The evidence-based guidelines the support this measure have not changed since the measure was approved for trial use and remain current.

Earlier this year we submitted this measure to the measures under consideration list for CMS in the map to review for potential inclusion in federal programs. In closing I would like to introduce Dr. John Wong who participated in the GI technical expert panel in the initial development of these two measures.

And he's also currently the co-chair of the PC (unintelligible) preventive technical expert panel. Dr. Wong serves as our clinical expert for these measures. Dr. Wong are you on and if so would you like to add anything?

Dr. John Wong: Thanks (Kerry), pleasure to participate in this. I just wanted to add a few brief remarks since we last submitted this measure. One is that when the CDC looked at and compared all causes of mortality due to reportable infections Hepatitis C alone exceeded all other causes of the 60 other reportable blood-borne or otherwise infectious diseases reportable within the United States. The other point I would like to mention is that compared to three years ago the entire landscape of Hepatitis C therapy has improved with once a day oral drugs that are taken from eight to 12 weeks with over 95% cure rates. And that obviously impact screening because it's clear that we have therapies that are highly efficacious and much more easily to adhere to and take for relatively short durations, happy to answer any other questions that may come up but I just wanted to acknowledge those two new developments since our last application.

Man: Thank you. With that I will pass it over to (Lindsey). If you would like to give us your brief summary, anything that hasn't already been said and then move us into a conversation and discussion around evidence.

(Lindsey Bachford): Sure. I mean I think for the sake of time I think this is a fairly although it's technically considered a new measure for endorsement probably a somewhat

familiar one to us given that it's a US Preventive Services Task Force recommendation. Although it's new official measure that was previously approved for trial use and so it's not also completely new to that committee. The - let me pull up this page here.

So I don't think - there was new evidence added since it was initially approved for trial use and I'll - instead of keeping a full summary I'll just go bit by bit so we can include the voting as well. So just with regard to evidence there was new guideline and again supports like the developer was talking about, I think really only strengthening the evidence and nothing that would call it into question. So I'm not going to speak anymore. I'm going to turn it over and we I do have some comments about the measure but I think let's make them maybe individually as we go through.

Man: (Don) do you have something? Any comments related to importance for measure looking at the evidence? All right anything else you would add (Lindsey)?

(Lindsey Bachford): Not at this time.

Man: Awesome okay. I think we can move to vote.

Woman: We will now vote on importance to measure the evidence on Measure 3059E. The options are high, moderate, low and insufficient. Waiting on two more votes. Just waiting on one more vote.

Woman: Give us - just give us one minute.

Woman: So the polling results ten committee members voted high, nine members voted moderate on the evidence criteria. So the measure passes on this criteria.

Man: Thank you. Performance gap?

(Lindsey Bachford): Sure so under performance gaps the preliminary rating is high. I agree with that rating. I think I will note when we talk about this measure there's two main groups that they look at or two data sets they use. One is Cerner data with over 800 providers. And then the E gaps or data is the other main samples that they use but 180 providers. The performance rates in both of those samples were quite low 20% and 34%. And those percentages are actually even less than what, you know, national suggestions of what performance rates are. So the gap is significant looking at multiple different data sources as well as nationally available data and so I think they have demonstrated a pretty significant performance gaps here.

Man: Thank you. Any additional comments related to the performance gap? Okay I think we can move to vote.

Woman: Voting of important measure the performance gap the options are high, moderate, low and insufficient. This is Measure 3059 one time screening for hepatitis C virus for patients at risk. Just waiting on one more vote on one more vote. So we have 11 standing committee members who voted high and eight standing committee members who voted moderate on performance gap of 3059. The measure passes on this criteria.

Man: Okay now we're move on to scientific acceptability beginning with reliability.

(Lindsey Bachford): Sure so reliability testing received a preliminary rating of insufficient and so I guess one question for developer would be if they have addressed those insufficiencies. Other than that I don't know. Depending on whether they have are not there would relate to whether we I guess should have the discussion.

Man: Yes just for the committee this goes back to sort of what they were talking about this morning. This is going to come up through all of the measures that we're looking at for the rest of the day to kind of clue people in to the conversation so with that we'll take it to the measure developers.

Man: So I think the issue is was the measure tested as specified? I would say it was since we're, you know, we did provide both at the ones listed off the (unintelligible) events. Additionally the care studies for this particular data set we were at the provider level and if my colleagues want to add anything to make that (unintelligible) decision.

(Jamie): This is (Jamie). So I'll jump in and add a couple of things and then if there's anyone else from our team back at the home office they can feel free to add too. But with respect to the specifications and the design the measures they're basically built so that we don't delineate, you know, different populations I think of that nature based on the care setting. And so I think that that is a bit of a discrepancy with respect to how we can apply the information within the (ETF) form. I know that we've had recent conversations with NQF staff about this because as you cite and per our understanding has changed more recently with this round of submission as opposed to how we've done it in the past.

With respect to the ten events I would say that that's not typically something that's done within the actual specification. It is something that would be more on the side of the implementation and sort of reporting programs would put that sort of requirement if there is a specific threshold of number of patients much like there is in - although this is not - the measure's not in TPT but emotional therapy TPT. So I'll see if there's anyone - anything else that you wanted to add or is there any other questions from the committee?

Woman: And so I would just comment for those of you that haven't had a chance to look at all the numbers that one of the concerns related to the liability this last (unintelligible) quality reporting event was lower than - was the one or more quality of events was .68 which would be below the threshold that would be considered reliable. However when it was above the ten quality reporting event the data was considered relevant. So the concern is that gap between one and ten reporting events that there is, you know, there's technically not in inclusion in the measure for it but it would be considered unreliable so that's (unintelligible).

And there was a difference between the two data sets. So even with ten or more reporting events the first data set had a reliability of .79 which was acceptable but on the lower end whereas the other one was quite good with .96.

Man: Any other questions or comments? Okay I think we can move to the vote.

Woman: The polls are now open for Measure 3059 one time training for hepatitis C virus for patients at risk. For the liability the options were high, moderate, low and insufficient. For reliability one committee member voted high, 13 committee members voted moderate, one committee member voted low and five members choose insufficient. That's over 60% so this measure passes on reliability.

Woman: Should we move on to validity?

Man: Yes I think what I'll do is cave to Sam to explain this.

Sam: Sure. A point of clarification, because this is a get data that includes both individual and group practice providers settings of necessity the only way that

we can consider this as a committee unless the data are - the announcements are clearly separate, you know, so the vote for reliability would be to accept this at the group practice level. So if we do approve this measure it can only pass reliability at the group practice level (unintelligible).

Sure, so the concern is that if they're not separate because for level reliability testing is a function of a - oh okay, are we good because I will gladly stop talking.

Man: Can I...

Sam: Sure.

Man: Sam on the document, the NQF document it says the level of analysis is just at the individual clinician, but you're telling me that it's at the group level also?

Sam: And I may have misspoke on this one then. If this one's just the individual clinician then you're (unintelligible) the concern for this…

((Crosstalk))

Man: And that's why I'm confused about why you it have an insufficient rating.

Sam: So it's because of the ten quality reporting events. And my apology. I've got ten measures in this - in my brain, my apologies. So is everyone clear then on what we just voted on? I want to make sure folks are clear because there be a nuance we get to later measures that are a little bit different. Clear as mud right?

((Crosstalk))

Sam: Oh, (Don) (unintelligible).

(Don): I've got a question that's sort of parrots all this and if you could just explain. When a risk assessment is required in order to do the testing how do we deal with whether or not the questions are asked and answered in order to get to that especially around substance use issues and (unintelligible)?

(Lindsey Bachford): Sorry could you repeat the question?

Sam: The measure is dependent on having asked. You know, if you don't ask questions in the right way if the individual isn't in the office to ask your panel may or may not ever get to the issue of testing. So we're not - we don't have a measure for did you ask right? We're assuming that they're asking or we don't care if they don't ask or what - how does that work?

(Lindsey Bachford): So the way that the measure is constructed I hope that this is going to get to the answer for your question. So the way the measure is constructed is that within the initial population of the measure there are specific criteria. We're going to look for the actual patient age, if they were born between, you know, each of the years, if they - if there was documentations on the history or if they - it was reported that they had a blood transfusion. All of these are individual data elements that would essentially be built in and either documented within the EHR or other through something like the patient report or if it were to be, you know, historical information like a diagnosis of some sort. Does that…

Sam: Well, it's not ...

(Lindsey Bachford): You don't so how you're going to pull that data.

Woman: She definitely has both.

Sam: I mean age and, you know, that's easy right? But this is predicated on other risk factors including intravenous drug use. And actually what was real puzzling to me and I did a lot of Googling just to see what people say but I always thought that frequent - having multiple sex partners was a risk factor, having been born in China with Taiwan was a risk factor. And so those aren't being considered so that's a separate question. But asking and getting information about intravenous drug use, multiple sexual partners and so forth, these are not necessarily easy to do especially when patients who aren't sitting in their office and have a trusting relationship with you.

Man: Yes the way I interpret it is and I don't like the way I interpret it so I will just start with that is that it's not so much a measure of did they, you know, accurately identify active or previous injection drug users. It's more if they had identified it, was this test delivered. At least that's how I'm understanding it. But I 100% hear you as a former injection drug user that those conversations like having that in there doesn't necessarily mean that all persons with a history of injection drug use are going to get this. But those who are identified as having a history should. It's sort of how I kind of broke it down in my…

(Lindsey Bachford): Yes I suppose the (unintelligible) that is not necessarily just the - I was just going to say yes, I was just going to say and it's not a screening but it's a (unintelligible) measure. But it's not a screening in the sense that there is a specific list of questions that are administered in order to gather this particular data. It is focusing on whether or not they have been identified for the provider in order to report of the measure and pull them in to perform that one time (unintelligible).

Man: Okay so that's very helpful. And what it does for me is tell me that the national goals for Hepatitis C screening will never be reached through this measure alone. But the other question just to clarify why isn't frequent multiple partners one - it's not in the guideline either. I read the guideline. It's just it's mentioned as a risk factor but it's not being suggested that we use it as an indication for screening in this measure.

Man: And I would add to that too the stuff that I read too just in HIV space we highly recommend -- I don't know if there's a guideline around it so please educate me if not -- but it has to do with male to male sexual encounter also being a big risk for Hepatitis C that there were folks in our population that that was the risk factor.

Man: And a mother child transmission obviously is hugely important and my understanding is that both for Hepatitis B and Hepatitis C you're born in certain parts of the world, especially in certain Asian countries. Your risk is actually quite high. So one might say anybody born in Taiwan within x time period should be screening. I just want - it seemed like these opportunities were missed (unintelligible).

Man: If I could chime in for a second. The CDC many years ago identified frequent sexual partners in fact those that had more than 50 sexual partners as being at higher risk for Hepatitis C. Subsequent studies investigated the likelihood of sexual transmission of hepatitis C and found actually a very low incidence among heterozygous couples. And consequently the frequent sexual contents was felt to be an association, not necessarily causal and a risk factor and consequently the CDC never recognized frequent sexual trends - sexual partners as a risk factor for Hepatitis C.

There are a number of other associations with Hepatitis C and depending on the organizations that are looking at the information they may or may not be identified as a risk factor. We tended to stay with what's consistent with the guidelines and a stronger evidence space for a casual inference as opposed to association.

(Lindsey Bachford): And I would tag along to that the citations they used to support it, the classes chosen are the ones that are in the ID Society of North America. They're the ones where the recommendations come from. So it's not that there is an increase list but in terms of again the strength of evidence they are using the ones that the national societies have identified as the ones you should be testing for.

Woman: So I know I'm driving people crazy today with this question but I still don't understand if you're using DHR data and some of this data is not always in structured data fields how are you going to pull these people who don't fall within your birth date range that have one of those other risk factors? How are you going to identify that denominator subset?

(Lindsey Bachford): It is going to be largely dependent upon whether or not that information is captured electronically. So there are (Sumed) codes. I'm trying to think, (Beth) feel free to jump in if you remember of the top of your head. But semi codes or line codes are ICD 10 codes that are going to be assigned that would essentially if they're found in the record to be pulled in in to meet that initial population or the denominator. And so it is very dependent upon having that there much like with many - with any other E measure.

And I don't know if it helps but it - they - the plan is to use structured data fields that aren't just on the problem list. So it could be social history but as long as there's a (Sumed) code associated with it, it's potentially used. So it's

not going to be all-inclusive but to their point earlier you wouldn't be penalized should you just free text in social history of IV drug use. They just will not appear in your denominator as someone that should be checked. So yes they are missed in an opportunity to remind them to be screened but it's not missed by not accounting for them in the measure.

Woman: And just to be sure that I understand the measure that is being proposed this would be across the board any physician who sees a patient twice, correct? It's not a particular position population correct?

(Lindsey Bachford): It's not necessarily a particular physician population. It is those who would essentially be caring for these patients.

Woman: Any physician who sees a patient born between the years 1945 and 1965 which is the only robust one that you're going to be able to pull from the EHR right, any physician seeing patients within those years…

(Lindsey Bachford): Who satisfy the specific types of (unintelligible)?

((Crosstalk))

Woman: Who sees a patient twice, correct?

(Lindsey Bachford): Who - it depends. So we have - it's two visits if they are specific types of office visits. If it is a preventive visit in nature then that would be a one visit requirement.

Man: So I think what she's asking my patient with heart disease who sees his cardiologist twice he sees me twice, he sees his neurologist twice who's - who

is going to be the individual that the clinician that's going to be evaluated in that? That's her question.

Man: Be sure to use your mics so that the folks on the phone can you can hear it.

Man: And if any one of those providers did the screening then it would be satisfied for anybody but if none of those providers then who…

Sam: Well I don't know if that would be correct because it would have to be pulled from their system that they were aware that that screening had occurred correct...

Man: Right.

Sam: ...because like if I - let's say I'm being treated in an infectious disease practice for something else they're unaware that I've been screened for Hep C, that happened in my primary care environment, there's no data link nor communication channel between those two spaces than the ID physician if they saw me could get dinged for that because they're unaware of it right? So…

Woman: And even if the patient said I'm now getting that test repeated. I already had it (unintelligible).

((Crosstalk))

Man: They have to get evidence in their EHR…

Woman: Well, yes I (unintelligible).

((Crosstalk))

Woman: (unintelligible) a great pitfall but she'd be, yes.

(Lindsey Bachford): So I believe that was in the measure specification we include guidance about if there is - if this test was performed outside of the reporting provider's system and how it - to effectively document that. So there is a one code that is - it's I'm sorry I'm trying to pull it up right now so that you - I can (unintelligible).

(Beth Bostrum): (Jamie) this is (Beth Bostrum) at the PTTI and I actually have the guidance up. So the guidance that we provide within the measure specification is that if the provider does not know the exact HTVR and a test performed elsewhere they can report the generic (unintelligible) HTV RNA panel. It's Code 75888-8. And it's found in the value site titled HCV RNA Test. And that's true. We provided guidance as well. We have it for HCVR in (Apac), they HC the antibody test and the HC the (repa) test.

Sam: So that would cover like folks who receive a rapid Hep C screening out in the community and can just report to their provider that this happened okay.

(Beth Bostrum): That's correct.

Man: Could I also just make one comment about the identification of at risk individuals? I recognize also as a primary care physician the challenges with identifying some of those subgroups. But I do want to make sure the group is aware that of the unidentified individuals with Hepatitis C, the CDC estimates that 2/3 of that group fall within the birth cohort between 1945 and 1965. So although the denominators for the other risk groups may be more problematic to identify clinically at encounters or even electronically, the measure applies

to at least what the CDC considers 2/3 of the individuals who are currently unaware of their infection.

Man: Any additional questions or comments? Oh I'm so sorry (Kim)?

(Kim): That's okay I get missed a lot. My only concern with it really was - well I should say my only concern. But my concern was that it really is not specific to a specific provider type. So it kind of flows fluidly. And that I think makes it a little more challenging to hold practitioners, specific practitioners accountable for the measurement and testing.

(Lindsey Bachford): What I'm hearing is some discussion that may relate to usability or feasibility later on. So maybe if we could get back to validity. So the developer used face validity on this. Obviously since (unintelligible) we don't have empirical validity and it does seem to meet the standard there. I think there are some questions exclusions for this one and for the other Hep C related measure. Patient exclusion is in there. I think that in some of the comments that was noted as maybe something that the group should discuss.

So patients can say they refuse to be tested and that actually counts as an exclusion. Now in real life that was not used that often. The proportion of providers who use that exception was relatively small so perhaps that's unrealized concern but yet a concern that was brought up in the pre-conversations.

(Kerry Fy): Hi. This is (Kerry) from the PCPI. I just want to clarify one thing regarding exclusions and exceptions for this measure. The only population that's excluded from this measure would people - would be people already diagnosed with Hepatitis C virus. They exception rates are a little bit different. Those patients aren't excluded from the measure.

The physician would use that patient reason if someone receiving refused to have testing done and exception rate for them reported back along with measure results so that they can see how many patients are - how many exceptions they're using. And those patients are still included in this performance denominator. So it's not like they're taken out like of actual exclusion would be. So I just wanted to make sure that were all on the same page with exclusions and exceptions.

Man: So would they - they would then be put in the numerator with the exception noted correct? No.

(Kerry Fy): It actually it would be a different rate. So they wouldn't meet the measure. So you wouldn't meet the measure.

Man: Okay.

(Kerry Fy): …but it would be additional information fed back.

Woman: Okay.

Woman: I guess I'm just trying to understand why is there an exception because if they're not included in the numerator or they are included in the numerator why would you have an exception at all?

(Lindsey Bachford): So the way that the exceptions work is that you essentially work your way through the measure so the initial population denominator and then you look to see if there are any exclusions, remove the exclusions. As (Kerry) mentioned that's the chronic Hepatitis C. Then we would look to see if the numerator was performed and if I wasn't before counting it as an actual

measure fail we would check to see if any of these applicable exceptions existed.

So if the patient absolutely refused to have this documented, if I - if they had any of the noted medical reasons in their history and their patient record they would be then (unintelligible). And so it doesn't mean that they could meet the numerator because by all means they could. It's just if the numerator is not met then we would look to see if these have been identified and then they would be removed from the numerator and from the denominator calculation.

Woman: I'm sorry a follow-up question to that. Are you measuring tests being ordered or tests being done?

(Lindsey Bachford): These are tests being performed.

Woman: So if a patient has blood test ordered in my office and they failed to go and get it done whose - that being the physician the physician takes the hit correct?

Man: So in - is there some sort of level that where the exceptions hit a certain number that you get concerned about it?

Let me put this in context. The reason I ask is because Hepatitis C, up until recently and still continuing still has a high level of stigmatization to it right? A lot of people believe that you only get it through injection drug use. And so there is some hesitation at least in the physicians I've spoken with and the infectious disease arena that this sort of practice I work in is that by offering it you offend the patient right? So there's a reluctant sometimes to offer and then the patient gets offended and then they say, "Oh, I don't need this test. I'm not that person."

And so if you have the ability to then, you know, say, "Well this is an exception, the patient declined," it allows you to sidestep a critical conversation right around stigmatization and patient self-stigmatization more than physician stigmatization and engaging the patient in that conversation about the importance of a Hep C screening and how it's not just injection drug use. So that's why we see this a lot in HIV. This is the same thing with folks like I don't want to offer that test. It creates an uncomfortable conversation. And I think Hep C kind of sits in that space a little bit.

(Lindsey Bachford): Yes, I mean I can definitely see where that's coming from. And I think that to speak to one other point before about the, you know, the number of visits that we require and things of that nature you know, one of the things is for these measures and other preventive measures that PCP I stewards is that we sometimes create a requirement so that there's two or more office visits.

And that is actually designed so that we can establish that there is actually a relationship between a patient and provider. So hopefully it sort of circumvents what you just described. However we do have the exceptions because there are certain instances that we can't really - we don't want to have the provider penalize inadvertently because of something that they - it's outside of their control necessarily.

But we typically include - we have these conversations with our technical expert panels or expert workgroups whichever you like to call them about, you know, whether or not these exceptions or exclusions are still applicable and appropriate for the measures. If there's information that we can provide showing that there's a very high number of patients who may be fall into one of these other categories maybe that's something that we would want to consider as an exclusion. Maybe if it's never being reported what's the point

in having an exception right? But we do want to be sure that there is a good reason to have an exception or an exclusion before we pull it into a measure.

Dr. John Wong: And this is John Wong. I just wanted to chime in about the stigmatization. I happen to have had the pleasure of or honor of participating in the CDC deliberations about the birth cohort 1945 to 1965. And that was an important factor in thinking about just of year of birth and being able as a physician to just say that the CDC recommends this not for any particular risk factor except for the fact that you're born between these particular years the CDC is recommending this testing.

Man: Okay.

Woman: And I just want to be really clear because when we said it before you said that they were kept in the numerator if they were an exception. But then I think I heard you say right at the end of that statement that if they were exceptions you would remove them from the numerator and the denominator and that wasn't what I read in the specs.

(Lindsey Bachford): The former is the latter wasn't what you expect?

Woman: Either.

(Lindsey Bachford): I believe the latter is part of the specifications where we have in the calculation algorithm section talking about how we go through. To (Kerry)'s point we do ask and recommend that there is an exception rate reported so that we know just what - and I we're talking about, you know, we do know how high that exception rate is if that is going to have significant concerns for the measure, if there are other items that we need to address within the specifications or other inclusions to consider.

Woman: So do they come out of the numerator and denominator or do they stay in?

(Lindsey Bachford): They come out of the numerator and denominator. (Kerry) please correct me if I'm wrong.

Man: I guess silence means you're right, right? (Unintelligible).

(Kerry Fy): For exceptions they stay in the denominator. They do not get counted in the numerator as meeting. And overall with the use of the exceptions I'll just add this one more small caveat, overall we've seen low rates of exception use. So I can understand concern of you think folks are going to mis-use the exceptions or potentially try to gain using them. We have not seen that overall for our measures that use exceptions.

We've seen fairly low exception rates within the testing data that we have currently. It's, you know, less than 1% in one data set and maybe 5% in the other data set that use the exceptions. So we're not seeing, you know, 35% of people with the patient reason for exception we just we don't see that so we don't see them used in that way.

Man: Any other questions or comments related to validity? Okay I think we can move to the vote.

Woman: Voting is now active for validity Measure Number 3059E one-time screening for Hepatitis C for patients at risk. The options are moderate, low and insufficient for validity. So for the validated criteria we have 15 committee members who voted moderate and five committee members voted low so the measure passes on this criteria.

Man: Thank you. Moving on to feasibility.

(Lindsey Bachford): Yes I think some of us started to kind of get in the conversation a little bit on feasibility earlier. The preliminary rating was for moderate and taking into account I guess some of the concerns earlier between the ability to capture especially those special groups in the EMR that need testing. But overall I think I would recommend that really I would say the rating for feasibility should remain moderate.

Man: Thank you. Any other questions or comments? And also we have NQF's E-measure team on the phone so if folks have specific questions related to that we have that resource with us as well.

Woman: Yes just a concern I have is kind of broad to the once ever types of measures. So I think it's hard to capture that. So somebody this requires that they have had this test done once and, you know, it could have been in ten years ago. And I think that those are hard to currently capture in our health records.

Man: I can tell you as a person who's leading a project trying to identify all those screens for Hep C and ever treated it has been a nightmare digging into the paper record pre-EHR so much so I've recruited an entire team of medical students to do it with the promise that they can have the data on the backend. But it has been weeks of work and that's just on a patient population of about 900.

Yes and I should back into it saying it was meaningful and I'm glad that we did it right? It was important information for us to know about the population of people we were serving but that it was a significant amount of work to find that information in paper charts in advance of the EHR being adopted into practice.

(Lindsey Bachford): I think to some extent that comment gets a little bit more towards usability. I completely agree with that. I mean feasible - and my interpretation is I mean, is the data in the EMR yes you can identify a birth date and then yes you can order a lab test or you can document history of HVC screening. They are - it is feasible to do it but the usable piece I think is really what you get to.

Man: Any other questions or comments about feasibility? The only other one I had was just a broad comment about broad comment about the likelihood of a person loving an ICD-10 code of injection drug use of their medical record. Like we have a lot of people that even hide certain diagnoses related to their healthcare. And active and past injection drug use is a pretty big thing to have on your problem list.

Woman: So actually that I guess is part of what I was saying about these things - we don't - we the physicians I think you need to understand we don't we very frequently don't put those in structured data fields for a really good reason right? So all of that in my history is pretexted in right? I don't put it in as diagnosis codes because it labels the patient. It has other implications.

Woman: One of the unintended consequences of naloxone standing orders for throughout the states is if you have a script you are prevented from getting life insurance now. We have many, many states. You have to be - that's probably why there's been a delay in putting diagnosis if you're still - because of your life insurance or you may lose your job because you had back surgery and you were on three weeks of opioids and they gave you naloxone to make sure you didn't overtake it or which is a lifesaving process and the same thing but you get an unintended consequences of not being - of getting dropped from your life insurance or health insurance.

Woman: So follow-up to that question for the measured developer, why did you not just take the birth date years?

(Lindsey Bachford): This is an option. This is not a combination. It's not used to meet multiple criteria. It is based on the guidelines, the USPSGS guidelines. And so this is the date, the birth date year. It's just one way in order to be pulled into it. The diagnosis of active injecting drug user also another option.

And I think that what you are - what's being described definitely makes sense and it's concerning in the sense that we're - what we can expect is that we wouldn't have as large of an initial population in the denominator as we likely should have based on how information is documented or how information is provided to the provider, you know, whether or not the patients feel comfortable speaking to that and having that assigned to their EHR record.

Man: I think all of us agree that the denominator is not going to include all populations that are recommended. We all agree on that. And I would have a real concern about that if the baseline rate of performance was 90%. But baseline rates I heard 20%, 38% things like that -- very, very low rates. So even for those that we can capture performance rates are extremely low. So I think I support the measure based on the huge opportunity for improvement.

Man: Can I also add that by having the knowledge that a particular patient even if it's undocumented within the codes, the clinician knows that the patient is a former injection person who injected drugs may prompt them to go ahead and screen because they are aware of this particular measure if it is approved.

Man: Any other comments on feasibility? Okay we can go ahead and move to the vote.

Woman: Voting is now active for feasibility Measure Number 3059E the options are high, moderate, low and insufficient. Voting results for feasibility of Measure 3059, three committee members voted high, four team members voted moderate and three committee members voted low. The measure passes on this criteria.

Man: And we blacked out the screen. So what we'll do now while our tech team is working diligently if we want to go ahead and move into use and usability.

(Lindsey Bachford): All right the next conversation is around use. And this is the pass or no pass measure. It talks about the extent it's being used now. It is currently in use in (MIPS) but no other feedback beyond that.

Man: Any other comments on use and usability? It's like no one has a comment while the screen's down. It was better. It's like no we're good now. So if you give us just a minute so we can get our screen back so we can cast our vote.

((Crosstalk))

Woman: I just have a comment. This is an accountability measure. And so as such it just raises some concerns to me if this is an accountability measure at the level of the physician that I think that would - I don't think that's fair to hold docs accountable for this when you just told us how difficult it was for you to even find that information.

Man: I would point out it is a voluntary measure though. You have to choose to report it.

((Crosstalk))

Woman: They're all voluntary at first.

Man: What's that?

Woman: They're all voluntary at first.

Man: Yes.

Woman: I'm sorry one point of clarification. This e-measure is not actually part of (MIPS) we just submitted it the registry version as part of (MIPS). It has been for the last two or three years. It's - I can't remember if it's 387 or 400. This measure was actually submitted to the mock list for consideration for 2021 though.

Man: And I would say too I think if I were to share with most of the other practices did to be honest they just redid the test on folks. And I just had a really big problem with that because there was a cost issue in healthcare. And I was like if the test has been done let's find it in the chart. But almost everybody else reissued the test is what they did in which, you know, comes to concern of paying for it right, like if people have insurance who can afford it or how much we're costing Medicaid as well.

(Lindsey Bachford): So for the developer clarify documentation of previous history of screening would count. It's not the actual result needed on the chart to say. You don't technically have to track down that report. You have to document that it was done on X date?

Woman: That's correct.

(Lindsey Bachford): So to your point still leaves the patient having to remember that they had…

((Crosstalk))

Man: Exactly yes.

(Lindsey Bachford): …(unintelligible) be able to affirmatively tell you that but…

Man: Yes and especially when going back into active drug use right like looking at a person, you know, tell me about all your screenings, you know, 20 years ago while you were shooting dope right? It get - it - because the recall is difficult. But I agree. I mean I think this is a good measure. I'm not trying to, you know, push it down in any way but just kind of pointing out some of the complications with it and that there is a significant amount of effort in collecting this information.

Woman: Just had a clarification. I think one of the requirements for measurements was the fact that patient was in hemodialysis that's right? So it's some (unintelligible) these measures want to (unintelligible) or do you have to continue those (unintelligible)?

(Lindsey Bachford): So I believe that that may be part of the initial population as another option. If when we talk about human dialysis it's going to be an overlapping exception of the measurement period. And so it's not necessarily something that is going on through the end of the measurement period. It's just that the patient is - seems to be actively proceeding from the dialysis. That doesn't help? That doesn't help?

Woman: I mean I understand how you framed it to get it one time. I just think that's the one time it's not sufficient in more population I think because it's an ongoing risk (unintelligible).

(Lindsey Bachford): One time so…

Woman: I think.

(Lindsey Bachford): Yes so I think in terms of how it's captured the hemodialysis is something that it could be something that's ongoing. It's likely something that's ongoing and is not given once ever. I think the patient's generally have multiple sessions of hemodialysis.

And so in order to be pulled into the measure specification they need to have gotten it again during the measurement period at some point. It doesn't necessarily say like absolutely one time right? We don't specify a certain period of time for it though just to make sure that the patient has received it because we - we're talking about looking at patients who are seeing the measurement period up through essentially December 31 of the measurement period if we're talking about a calendar year but also for our patients who are seen January 1. And so if we know that they're on hemodialysis on January 1 then that would satisfy the requirement.

Man: Any additional comments or questions? All right we can go ahead and move to the vote.

Woman: Voting is now active for the use criteria on Measure 3059E the options are pass and no pass. So we have 19 votes in, 17 committee members choose pass and ten members choose no pass.

(Lindsey Bachford): So the final question for us relates to usability. And the two main parts are have we shown improvements and then did the overall benefits outweigh the harms. We - hasn't been used so we don't have any scores for the first part so really the discussion should be around the second one is do the benefits outweigh the harms? And the harms identified were that period to getting your test results.

And then I think the only additional one added by Adam over here was the unintended consequence of additional testing to prove that a measure has been met or over testing to prove the measure has been met. I think they are real things but the potential that huge gap to close is still outweighed by those potential harms. So I would agree with the recommendation of moderate in terms of overall usability.

Man: Thank you. Any additional comments? Okay we can move to the vote.

Woman: Committee members have the ability to choose high, moderate, low or insufficient on the usability criteria for Measure Number 3059E. Waiting on one more vote. One committee member chose high, six team members choose moderate, one committee member choose low and one selected insufficient for usability. The measure passes on the usability criteria.

Man: Thank you. And there were no related or competing measures and no public comment so with that we would move to suitability for endorsement.

Woman: I think we've had plenty of discussion and can probably just vote.

Woman Overall suitability for endorsements committee members have the ability to choose yes or no. On overall suitability on - for endorsement on Measure

3059E One Time Screening for Hepatitis C virus or Patients at Risk 16 committee members voted yes and three members voted no.

Man: Thank you very much. With that I will thank the lead discussants from this morning and this afternoon for all of your support as well good discussions and I'll happily turn it over to Dale at 2:45 for the rest of the afternoon.

Dale Bratzler: So Adam and I last, we had a conference call a prep call and we said do you want the morning or afternoon, you know, Adam chose morning so. All right we'll move on to Measure 3060E which is Hepatitis C Testing in Patients with Ongoing - I have to use the exact term here - with Screening for Active Injection Drug Users. So a lot of the concerns that we've had particularly around defining the denominator population probably be even more prevalent for this particular measure. Our lead discussant was (Tim Elliot) (unintelligible). So I guess do you have any other comments about this particular (unintelligible).

(Kim): (Kerry) or Dr. Wong anything else about this Measure 3060E?

(Kerry): Hi.

Dr. John Wong: Hi Sam. I just wanted to highlight how infectious hepatitis C is among injection, people who inject drugs. There are reports that within a year 80% of such individuals may acquire hepatitis C. Other reports cite somewhere between roughly about a 20% risk over that first year of use of injection drugs. I'll also point out that 85% of people who inject drugs do so with other individuals on average for others and they commonly will share either syringes, needles or other drug use associate paraphernalia. And up until about 15 years ago the prevalence of hepatitis C reliably decreased from older ages where it was mostly blood transfusion transmitted from what we knew as non-

A, non-B hepatitis to increasing incidence and prevalence in younger individuals in their 20s that are described generally white with equal men and women and mostly in rural areas.

And I'll just point out in a study looking at trying to reduce the sharing of syringes and needles roughly about 1/2 of them were doing it because it was a social norm in their group despite knowledge that they could be increasing their risk. I'll also highlight the fact that 72% of these younger individuals with hepatitis C were unaware of their, who actually had hepatitis C were unaware of their status and that, unaware of testing. And that even those young people who injected drugs who were not hepatitis C positive 50% of them were unaware of their status and consequently had not been tested.

Dale Bratzler: Okay, thank you. (Kim)?

(Kim): This is very similar from every section to what we just talked about in the last measure. But this is Measure 3060E Annual Hepatitis C Virus Screening for Patients Who are Active Injection Drug Users. And this is also a PCPI measure. And the description as a percentage of patients regardless of age who are active injection drug users who received screening for HCV infection within the 12 month reporting period. So it is an annual measurement. And it is at the clinician and individual level.

And talking about evidence this is almost identical to what we talked about in the last measure. It has the same clinical guidelines with the physician of -well I think that was in the last one to the US Preventive Services Task Force. Clinical guideline is the AASLD-IDSA from 201. And that was graded to Level C which is a consensus of opinions of experts, et cetera. And (unintelligible) that it was moderate quality evidence. And it had the same clinical studies the Cerner and same medical record system essentially there -

that they tested this particular measure on. And it had similar levels of results. Does anybody have any questions about the evidence?

Dale Bratzler: Does anyone have any other questions about the evidence for doing tested in active injection drug users? (Don)?

(Don): No, it's not really a question about the evidence just something to put out there as we consider this and not everything desirable needs to have a measure. You know, we are trying to reduce the number of measures. And I would have hoped that a provider who is done initial screening and has a patient with intravenous drug use in front of him or her would continue to monitor the situation. So I'm not sure why we need to have the system do another measure not that it's not important I just don't - this might be an opportunity to be smart about what we measure.

Dale Bratzler: What is the rate of the recurrence after treatment, you know, with Harvoni or Simponi one of those…

(Kim): That data was not in what I read about recurrence.

Dr. John Wong: It's roughly about out of 100 individuals two the six of them are the most widely cited numbers. As I mentioned there are now single pill treatments. And again response rates exceed 90% even in this population. And indeed there is even a salvage regimen that contains free drugs for individuals who may develop resistance to it. I just will point out again that the risk in this group is around 20% per year climbing up to 50% by five years.

And again some studies suggest even higher rates than that depending on the population being studied. And so in distinction to the prior manager this is similar in some sense to hemodialysis but carries an even higher risk. And

from a population health study there is this (transmitability) among people who are recognized as injection drug, people who inject drugs and this tendency for sharing of needles and paraphernalia among at least in this one particular study among four other individuals.

(Kim): Some other data that they cited from studies was that African-Americans were more likely to be screened in Caucasians and men were more likely to be screened than women. And they also talked about this particular issue being epidemic level in the Appalachian region of Kentucky, Tennessee, Virginia and West Virginia.

Man: I'll add that it's also epidemic in Massachusetts, New York and Wisconsin.

(Kim): It wasn't in what you wrote but okay.

Dale Bratzler: All right any other questions or comments about evidence? (Don), I think you're okay. (Don)'s not even there. All right, we can go ahead and vote.

(Asaba Denwulquafor): The poll is now active for evidence on measure 3060(e), unwell Hepatitis C screening for patients who are active injection drug users. We're - options here are high, moderate, low, and insufficient. Wait for two more votes.

So, for evidence on 3060(e), we have 14 committee members voted moderate, four committee members voted high, one committee member voted insufficient. So, the measure passes on this criteria.

Dale Bratzler: Go ahead, (Kim).

(Kim): Okay. So, in relation to GAPs, they didn't - the developer didn't provide any independent disparity data. They decided - they cited the articles and studies related to Caucasians and women being less likely to be tested. Also, discussed that former and older drug users are more likely to be infected.

That - those were the only disparities that they really addressed. But there definitely is a gap in the number of people that are being tested and the number of people that probably should be tested, based on the data.

Dale Bratzler: Okay. Any other comments about GAP? I think we can go ahead and open the poll.

(Asaba Denwulquafor): So, the poll is now open for GAP. The committee members can either choose high, moderate, low, or insufficient. We're voting on measure 3060(e), unwell Hepatitis C screening for patients who are active injection drug users.

So, of the 19 committee members who voted, 11 chose high, seven committee members chose moderate, and one committee member chose insufficient. That's on performance gap measure 3060(e), and it passes on this criteria.

Dale Bratzler: We'll move to the conversation about reliability now, and the same issues that I think you have discussed for the last measure apply to this measure, to the inability to test blow the threshold of ten or so.

Man: Right, and my apologies to the measure developer for previously conflating this with other measures. The other thing that we wanted to point out is that, in staff review, there was a little bit of confusion as to whether or not there was representatives inside of the sample for care setting.

So, when measures are specified by care setting as well, not just by level of analysis by which we mean the type of measured entity, but also with - that was actually very clear from the submission. But the issue that we're bumping up against was one related to care setting.

Now, this has been part of NQS measure evaluation criteria, so it's not new. However, we do want to be clear that there has been an added emphasis on some aspects of how we have the expectation for adherence to NQS criteria.

This isn't one that we've - the term that gets used often is "die on our sword," on - where we make it so onerous for measure developers to comply to the NQS evaluation criteria, that it becomes overly burdensome for them to accomplish it.

And we intend to allow for a runway of time for which we - well, I'll shy away from saying enforce our measure evaluation criteria, but where the standard is known and expected and where our scientific methods panel or our CSEC might make a very strong recommendation that we adhere to it.

So, this isn't one where that's the case, but it's something that may be helpful for the committee, if the measure developer provides some clarifications around which, precisely, of the specified care settings were represented in your testing sample.

(Sam Tierney): Yes, hi. This is (Sam Tierney) with the PCTI. Can you hear me?

Woman: Yes.

(Sam Tierney): Okay. Okay, great, thank you. I just wanted to - before we get to your question, I just wanted to clarify something, a source of possible confusion, I

think, related to the reliability testing. So, we did test the measure as specified. I think (Jaime) spoke to that earlier, which I know is an NQS requirement.

And for - I would say for purposes of expanding upon the reliability testing, we have historically provided the reliability testing at one event and ten events. And so, that's what we did here. I think it was potentially confusing the NQS GAP, as to why we did that.

It's something we've historically done, because I think there's an interest in knowing, you know, as the number of events increases, does the reliability increases, and is there a number at which it's not reliable? So - but we did provide both of the results within the testing data, so you have that available for you to review.

As it relates to the care settings piece, and I guess I would say, you know, as a measure developer, you know, we really very closely try to adhere to NQS requirements and as they're described. I think additional clarity has been provided as of late.

That's not necessarily well spelled out in the document, so I know we really look forward to seeing that specificity added, so that we can, you know, know what we're aiming for and the target that we're trying to achieve.

But I will say, you know, the care settings piece, I know you had a specific question about which care settings we believe it was tested in. I know - I don't know, (JR), if you could speak to the testing data on that. But I just wanted to offer those additional clarifications as it relates to the testing information and how we executed on that.

(JR): Yes, so if you look at the data set two, the Capricorn data set that's on - in the submission, that data's comprised of multiple hospitals and also multiple sites, which could - it was an epic representation, and also a GE centricity.

When they do import those records into their data warehouse, the Capricorn data warehouse, they're - you lose some of the specificity as to what care settings are represented.

So, we did request the site to provide a little bit more detail, but they said they could not provide that. So, for - depending on where you get data from, you might lose some of that detail, so just something to think about.

(Sam Tierney): And this is Sam again. Just to clarify, (Jaime) mentioned this earlier, but there is no different specification per care setting. It's just the one specification, and then we have the various coding included that might lend itself to different care settings, based on that particular code. So, I don't know, (Beth), if you can come in on which care settings are included currently in the measure, just to give the committee a sense of that.

(Beth Foster): Yes, Sam. Thank you. This is (Beth Foster), from the PCTI, and then within the initial population requirements for the measure, we have specified that a patient needs to have one visit - one preventive visit during the measurement period, and then - can you all hear me okay, or am I a bit quiet?

(JR): We can hear you.

(Beth Foster): Okay, wonderful. I'll try and speak clearly. But like I mentioned, we have a requirement of two (unintelligible) two visits during the measurement period, or one preventive visit during the measurement period. So, within the measure

specifications for the ECQM, we've identified on the following care settings as appropriate, as having, you know, part of the two-visit requirement.

And the value sets are as follows. Its care services in a long-term residential facility, nursing facility visit, home healthcare services, outpatient consultation, an office visit, and we've identified codes that are appropriate for a face to face interaction.

Then, the preventive visit that we've identified as appropriate are annual wellness visit, we have preventive care services with established office visits for those 18 and up, preventive care services for the initial office visit for 18 and up.

And then we have preventive care services for those initial and established from zero to 17, and the way we've identified those values that's encoding as appropriate is we worked with our technical expert panel to identify those values as encodes, and ensure that they be appropriate for measure inclusion.

Dale Bratzler: Okay, thank you. So, any other conversations about reliability? Remember on the last measure, we did move it forward on reliability, similar testing set.

(Beth Foster): So, I guess maybe it's not exactly reliability, although the data was presented in that section, so I'll comment on it here. The occurrence rate was so small.

I mean, out of - in data set one, there were only 30 that ended up being recommended, and then data set two, although 22,000 sounds like a lot, you know, that was in over 4.8 million visits, so I just kind of wondered that population we're getting at is so small, I think it's due to the issue of who's self-reporting as an active IV injection drug user. It's a pretty self-selecting population.

If they're going to admit it and be okay with putting it on your health record, they're probably going to be the ones that are going to be okay with getting tested then, too. So, again, maybe it gets more towards usability, but the numbers in the reliability testing were shockingly very small for me.

(Kim): And I would say one other thing about that, too. The ones that were included were providers that had ten or more qualifying visits. So, anyone that had one, two, three, up to nine were not included in the sample. Just for consideration.

Man: I would also say, at least in my experience, most active injection drug users, myself included at the time, our care is unscheduled, right? We're not scheduling office visits to come in, and so I think the exclusion of, like, ED's and those kinds of locations, acute care centers where people are likely - who are actively injecting, to get that kind of care, means you're not going to hit that population, because that's not very accessible.

Dale Bratzler: Any other comments about reliability?

Woman: I have just one more question. So, are ED's excluded from this measure?

(Kim): Yes.

(Carrie): Hi, this is (Carrie). The ER setting is not currently included. It doesn't mean that we couldn't consider adding it, but the - it presents some of its own challenges as well. I'm not saying that you shouldn't do it, but there's other things that need to be thought through there as well.

So, you know, then you get the tests done, then you have follow up and getting those folks connected with follow up care. So, I think that's something

that was probably once discussed, and it requires us taking it back to a technical expert panel and further fleshing it out. So, it wouldn't be adverse to adding it, but it's not in the measure as it's currently specified.

Man: I would consider it.

(Kim): Well, and NQS said they scored the reliability as insufficient.

(John): So, I have my third recommendation of the day, (Frank), in addition to the other two, which I'm sure you remember vividly. There'll be a test later. But I think it'd be good to ask developers to provide the attributable fractions of what they're doing for the population-based problem.

So, if they're - just to put a number out there, 200,000 people at risk for Hepatitis C or intravenous drug users, what is the attributable fraction for this measure driving improvement?

And as was the case for the one we discussed before, it's probably really low, given the - where the problem actually is. So, just - these are things, I think, since I'm not on the NQS board, I'll make these recommendations now. I think that would be a good change to the procedure.

Dale Bratzler: Any other comments? Did you want to respond to that?

Man: I'm happy to, (Kim), if you'd like me to.

(Kim): Please do.

Man: Alright, so NQS staff's perspective on this is that we want to establish what the criteria are, and then take an approach to presenting that evaluation in a

way that allows for some flexibility for the committee, in the event that that's available.

So in this instance, the reason that we've provided the clarification that it's not something where we're dying on our sword is that we're not going to require that this be evaluated in a certain way.

And - but I wanted to make it clear what the expectation is, and in the future, it can very much become a reality where we get a strong recommendation from CSEC or from the Scientific Methods Panel to adhere closely to this. But at this point, that's not the case. So, we'll leave it to the committee to determine how strongly they wish to adhere to the NQS criteria.

(Sam Tierney): And this is (Sam Tierney), from the PCTI. Could I just ask (unintelligible) staff to clarify? My understanding of the source of your insufficient rating was this kind of confusion over the one in ten event.

But if you look at the results for one event, you know, because understanding that we haven't specified it to have ten or more events, but if you look at the reliability and the committee looks at the reliability for ten events, those would be - I mean, for one event.

Sorry, I misspoke, those would be the numbers that I would think you would be basing this decision on. And I think - I wonder if you could offer some more clarity around that, because I feel confused, and I imagine the committee may as well.

Man: Thank you for the question, and so the clarification point, as I understand it, is whether or not having one or more eligible events is the appropriate thing to consider when it's been specified that way, and absolutely it is.

So - and perhaps the measure developer can clarify if, in data set two, which in staff's estimation was the one that was more critical, given that there was fewer than, you know - there was only a small number of qualifying reporting events in the first one, which kind of called some questions around the reliability analysis in general.

But with data set two, there being a lot more qualifying events, was there an analysis performed that was just at the specified level of one or more?

(John): So for data set two, it was assessed at the tenth level, reason being the site that provided the data had mentioned there was some risk in re-identification, if we went under the ten-plus events. So, if we - if they provided one-plus, they had some issues with re-identifying both patients.

Man: Understood. So, the expectation is that this be the testing aligned with specification, again. So, this is another example of something where we're not dying on our sword yet.

But - so - but just have - would like the measure developer to be under advisement. In the future, it'd be much better if you either specify your measure to have the exclusions or to test directly to the specifications in the second data set.

Dale Bratzler: Any other questions about that, or any other comments about reliability?

(Kim): I think just probably add some of the pre-evaluation comments, and there was overjoy about the use of standard coding for this particular measure, so that was good.

However, there was concern, again, about identifying all of the people in that population and readily identifying who was an IV drug user, so that will potentially be a challenge for this measure, in getting accurate and reliable results.

Dale Bratzler: Alright, are we ready to vote on reliability?

(Asaba Denwulquafor): Poll is now active for scientific accessibility of measure property through reliability, measure 3060(e), unwell Hepatitis C screenings for patients who are active drug users. The options are high, moderate, low, or insufficient.

Of the 19 committee members who voted, eight voted moderate, nine voted low, and two selected insufficient for reliability. So, that's 42 percent for moderate, 47 percent for low, and 11 percent for insufficient.

Woman: So, that's consensus not reached, because it's within that 40 to 60 percent zone. So, we will continue voting up to, but will not make a recommendation on overall endorsement. We will have to re-discuss at the post-comment call. Any questions?

Dale Bratzler: Alright, (Kim)?

(Kim): (unintelligible) this one, similar to the last measure, was face validity testing. They were unable to do empirical validity, since the measure's not in widespread use, in order to (unintelligible) required for the initial endorsement.

They had an expert panel that has 23 members, and it was a wide range of specialty provider types. And the specialty provider types on the expert panel

for this particular area is high, at 4.74. They also discussed what's in the last measure, about exclusion and exception analysis, and reported very low use of the exceptions in their example cases.

And as some of the pre-committee pre-evaluation comments, there was one person that identified the face validity as a potential concern, and the ability to identify intravenous drug users.

The developer did not include any results or analysis (unintelligible) ER during the measurement period, so that adjusted to the ER not being included.

It was not tested across different care settings, and that was thought to be a threat to the validity of the measure results, and the transparency of the measure by using - the use of the potential exceptions, even though their study showed very limited use of exceptions.

And the NQS staff were concerned by the high number of exclusions from the testing sample, and just communicating concerns about the 5263 exclusion. That was something that they thought we should discuss. NQS's board (unintelligible) moderate.

Dale Bratzler: Other comments on validity testing? If not, I guess we'll go ahead and vote.

(Asaba Denwulquafor): The poll is now active for validity criteria on measure 3060(e), unwell Hepatitis C screening for patients who are active injection drug users. The options here are moderate, low, or insufficient. We have recorded 19 votes from the committee members. 12 committee members voted moderate, and seven committee members voted low for validity. So, for validity, the measure passes.

Woman: That's 63 percent, depending (unintelligible).

(Asaba Denwulquafor): 63 percent for moderate and 37 percent for low.

Dale Bratzler: Alright, feasibility.

(Kim): Feasibility, so the data source is the EHR measure. There are 22 different data elements included in the measure specification. It is - the diagnosis has to be injection drug use, which was a low scoring domain for this particular review. It is not a widely adopted measure right now.

It does not require the clinician to do any additional data entry for the quality measure. It should be a byproduct of routine care for the patient. The data source may not be the most authoritative source patient, and it's not a complex measure.

One thing I did want to point out is that for this to become a measure in the commercial domain, that the users would need to have a licensed agreement with TCPI Foundation to be able to use the measure specs.

For feasibility, the pre-evaluation comments indicated that the data may be incomplete just because it's so hard to identify who the IV drug users are, so you may not have that complete and comprehensive group of individuals.

It may be hard to identify them in the EHR, depending on where that information is located, and I think that Adam addressed that briefly before, the challenges with that.

But the data is available in the EHR, as long as the patient discloses and that conversation does occur with the provider, and I believe that was - it's for pre-comment.

Dale Bratzler: Any other comments about feasibility, similar issues to discuss?

Man: Yes, one question I do have, and it kind of came up before, but I think I'm a little bit more concerned about it here. How do people make a decision between active and inactive for injection drug use?

(Kim): From reading the specs, it appeared it was patient disclosure.

Man: But I mean, as far as - like, I'm a former injection drug user. Like, what is the time period that - is there an accepted time period that somebody would move from active to inactive? Because I'm not aware of it, and it would be nice to know, because I think it's still my problem. I would love for that to switch.

(Kim): It wasn't specified in the measure, so only the developer would be able to say how they identified that.

(Beth Foster): Yes, absolutely. This is (Beth Foster) from the PCTI, so the way that we identified it in the measure specification is that the patient must be an active drug user at some point during the measurement period. So, in technical terms, it overlaps, but essentially at some point, they have to be documented as an active injection drug user.

Man: So in the last 12 months, you said?

(Kim): Yes, the measurement period is annual.

Man: So you would have to go in and change it, right, from active drug user to inactive.

(Beth Foster): Correct.

Man: Right, so there's a code for active, and we're saying you would have to go in and, like, switch it to inactive.

(Beth Foster): Yes, it's all - the way that it's specified, it's based sort of on the prevalence period, if you will. And so, there's a start time, and then there's an abatement time, or if not, then it is still an ongoing diagnosis. And so, if it were not indicated to have stopped previously, then it would still be considered active for the purposes of measure.

Man: Okay.

Dale Bratzler: Which is hugely problematic, in a problem with electronic medical records.

Woman: Yes, I think feasibility is where I struggle most with this one. I just am not really sure the population is the one that's coming in for two visits with the provider in a year and coming for a preventive visit.

And I just question whether this is - I mean, you're asking physicians to then document on a problem (unintelligible) history something that doesn't have a widely accepted definition, active versus history of.

You know, if they finish but they still have used within the last year, should they - I mean, they really should be tested, but they're not going to appear, because you don't want to add to their problem list saying active drug use,

then just to take it off and show that they've stopped. So, I just don't know how feasible this one is, and it ranks pretty low in my book.

(Beth Foster): I will just say that we have a definition within the measure to hopefully address some of the active versus history or inactive, that says that for the purposes of the measure, the active injection drug users are those who've injected any drugs within the 12 month reporting period.

And so, there is sort of that parameter because we're checking to make a numerator action is conducted during the measurement period as well.

And so, without being overly burdensome or overly complicating the measure to say that this has to all happen and be documented during a very specific encounter, because we recognize that, you know, we might be performing the actual tests, but the results may not be there, or they might be ordered during a specific encounter, but it might not be performed until maybe a week later.

There's sort of room for that by not tying all of this to a specific encounter and broadening it, so that it's applicable to the measurement period.

Dale Bratzler: (John)?

(John): I'm not sure if this is feasibility or it should go under use and usability, but I'd like to ask the developer why they felt this warranted a separate measure from 3059(e), given that 3059(e) will capture the same population. It's the same level of analysis.

Woman: The difference, I think - and I'll ask if my colleagues feel that they need to jump in as well, but the difference being that 3059(e) is actually a one-time ever for a broader population, whereas this is looking for an annual screening

for a subset of that, for patients who are at continuous risk. Thank you, (Elvia), (unintelligible).

Dale Bratzler: Any other comments about feasibility? Alright, carrying on, we'll go ahead and vote.

(Asaba Denwulquafor): Voting is now active for feasibility on measure 3060(e), unwell Hepatitis C screening for patients who are active drug users. The community has options to either select high, moderate, low, or insufficient. Of the 19 committee members who voted, four committee members selected moderate and 15 members selected low for feasibility.

Woman: This is not a must-have criteria, so we can move forward.

Dale Bratzler: Alright, (Kim), you want to talk about use and usability?

(Kim): Well, for use, the measure is not currently implemented in any of the performance programs, so there really isn't a lot to say yet about use. And QS did indicate that it is - that they would give it a pass for use.

(Jaime): And this is (Jaime), just to jump in, similar to the other measure. The registry version is included in QPP, and was submitted as well for the month list for 2021.

Man: So I only had one comment, and I think it kind of builds off what (Don) was saying a little bit earlier, which is that the only thing I could see as far as an unintended consequence is that if it's not measuring the environments where active injection drug users are likely to go, and an institution is scoring high on this measure, it could lead to a false sense that they're adequately serving this population when in reality, the denominator isn't including, potentially, a

large portion of the population by not measuring the environments where active injection drug users are seeking care.

So, again, it's not that I think it's a bad measure, just I think there is this potential without these other elements, and I know that's not what your measure's looking at, right, which is the, you know, screening people for active injection drug use.

But I just think it's an unintended consequence, that somebody could look at this performance and feel better about what they're doing without necessarily digging into it too much, to see if they're actually adequately identifying the population.

Dale Bratzler: I'm going to separate the conversation on use and usability, because use, currently it's not in use. So, NQS staff basically said, "We'll give it a pass on this. It's a new measure," essentially. Usability, I think, though, is where you're at, so let's go ahead and vote on use.

(Asaba Denwulquafor): Poll is now active for use on measure number 3060(e), unwell Hepatitis C screening for patients who are active drug users. Committee members can select pass or no pass. We're just waiting for one more vote.

Dale Bratzler: Taking us down to 18.

Woman: Okay, we're at 18 now. Okay, so we have 12 for pass and six for no pass, and we'll just convert this to percentage, but this is not enough pass criteria for a new measure. Okay, so we have 67 percent gave it a pass and 33 percent gave it a no pass, so it does pass, and we can move forward with more discussion on usability.

Dale Bratzler: Usability is next. Adam's comments, I think, were very much to the issue of usability and setting the care. Are there other comments about usability?

(Kim): Again, they didn't have much in the documentation that we received for this measure, because it hadn't been implemented widely. However, pre-evaluation comments from the committee said that if it was implemented consistently across multiple provider sites, it could improve healthcare outcomes.

The benefits for testing are strong, and the measurement should increase the focus on testing. So, that's something to consider, and there are no real harms, but identification of the population that needs screening will be difficult. And NQS rating, preliminary rating, was moderate for usability.

Dale Bratzler: Any other comments? If not, we'll go ahead and vote.

(Asaba Denwulquafor): Poll is now active for usability. Committee members can select high, moderate, low, or insufficient, and this is for measure number 3060(e). Of the 18 committee members who have voted, eight people selected moderate, and ten committee members selected low for usability. So, we have 56 percent low and 44 percent vote on moderate for usability. This is for measure number 3060(e).

Woman: Okay, and again with this one, it's not a must pass criterion, so - but we will not be voting on overall endorsement, because we did not reach consensus on validity, so we won't do that vote.

Dale Bratzler: There were no public comments or other related or competing measures, so I think that completes the evaluation of this particular measure. It will come

back to us on a post-committee conference call in the future, so we'll be touching base on this one again.

We're at a decision point right now. We have what - well, it depends on do we want to push to 4:30? I mean, we actually, on the agenda, originally were scheduled to stop at 4:00, for member comments.

Okay, so we're going to try to get through one more performance measure, because otherwise we're going to spend a lot of time on the phone doing these performance metrics that we didn't complete in the face to face meeting, we'll be doing by conference call.

So, our next performance measure is 0086, on primary open angle glaucoma optic nerve evaluations, and our lead discussant is (Scott Perek). PCPI? Yes.

Dr. (Scott Perek): Hello, this is (Dr. Perek). I'm a glaucoma specialist in private practice in Pennsylvania, and I've been working with the PCPI team on this measure. Can everyone hear me okay?

Dale Bratzler: Yes, very well.

Dr. (Scott Perek): Okay, good. Thank you so much for allowing me to do this on the phone. I'm in the middle of a busy clinic, and it's hard to get away for the entire day. So, I'm going to give a little bit of background about glaucoma and the measure, and then I have to of course get back to patients, and I'll turn it over to the team to go over the remaining points.

So, if there's anything that I say that spurs any questions in your mind, please let me know right away, so we can address them before I have to leave the

call. So, as everyone knows, glaucoma is a very important, very prevalent disease in our society and in the world.

There's some two, three million people in the US alone who have glaucoma, and that number is going to balloon to probably seven million in 2050. It happens more as we get older. The drainage system within the eye, you know, peters out as an age-related disease, and it can lead to permanent, irreversible blindness.

It takes away your peripheral vision first and your central vision later in the disease, and so unfortunately, people don't realize they have glaucoma until it's too late. And again, it's nerve damage. It's the optic nerve that gets damaged and destroyed, and so it's a permanent type of vision loss.

Obviously, this kind of vision loss leads to tremendous financial burden in our country, and again, around the world. It's estimated that it costs about $35 billion in 2004, and some of that is direct medical costs and then some of that is of course other direct costs, and of course, a large portion also is productivity losses.

The - glaucoma is tricky to detect, and you know, the - when the vision loss is permanent, it can lead to depression and diabetes and hearing impairment, stroke, falls, cognitive decline, and even premature death. Decreased ability to see can lead to difficulties driving and people losing their license because of failing the peripheral vision test at the driver's bureau.

It prevents you from being able to read and keep accounts and travel in unfamiliar places, so really, the burden is tremendous, and whatever we can do to catch this disease early before it's led to all these optic nerve damage and visual loss and cost is a good thing.

So, we have a measure, which is an examination of the optic nerve, which is maybe the most important factor in detecting glaucoma. Many of you know that eye pressure is a key component of glaucoma, but really, it's certainly a risk factor, and it's the only modifiable risk factor.

Obviously you can't change your age or your genetics or your race or anything, so it's a modifiable risk factor. But approximately half of people with glaucoma actually have an eye pressure that's within "normal limits."

You know, average and maybe one or two standard deviations, so using the eye pressure alone to screen for glaucoma, to detect glaucoma, you'll catch a bunch and you'll miss a ton. And really, the optic nerve is the key gold standard of detecting glaucoma, so there's many ways to look at the nerve.

One is just direct examination, using the float lamp, which is our, you know, routine way that we do it in the office, and there's also computerized scanning of the nerve and the nerve fiber layers, which gives valuable structural information about how much damage has occurred and how severe it is, and give us a chance to get a sense of whether things are progressing or not.

We have the American Academy of Ophthalmology as our professional society, and of course there's also an American Glaucoma Society, and the two work very closely together. They have come up with official guidelines.

We call them the preferred practice patterns, and it very clearly spells out that the optic nerve exam is a key part of any eye exam, especially in glaucoma, and it - the appearance of the nerve has to be documented year after year or, you know, month after month, however - whatever the frequency of the visits are.

And it has a strong recommendation with a very excellent level of evidence, and many well-conducted meta analyses and systematic reviews have been done, which demonstrate excellent, strong, and longstanding evidence of the importance of this measure.

Our new measure - or our measure does have a little bit of new wording, which captures, you know, all the same key points. There are certainly gaps in care, despite all of us knowing how important the optic nerve exam is.

There have been documented gaps in care in the literature, and also, you know, more recently as this has become a measure, we're able to use registry data, and we can see that not everyone is following the preferred practice pattern. And so, this is a very important measure for us to keep and keep it going.

The importance of this cannot be overstated. There are also interesting disparities within socioeconomic status, so Medicaid data has been studied and we can see that the number of glaucoma patients getting optic nerve exams in that population is lower than in the privately insured population.

There's also racial disparities, and as you know, Black Americans are at the highest risk of developing open angle glaucoma. So again, the gaps in care are important for us to realize that they're there, to address them, and also to study how it's affecting various populations.

With that, I'm going to close my comments, and again, if there's questions for me, I'll happily take them, and then I'll turn it over to the PCPI team to go over some of the other details.

Dale Bratzler: Okay, are there any other? We need to keep moving, if we can, so other details? Any at all? So, (Scott), do you want to start your conversation?

(Scott): Okay, so now we're all experts on glaucoma. So, not to reiterate what you said, so it's glaucoma - screening the nerve is very important in glaucoma. This is a measure that's previously been endorsed, up for maintenance, and there's really not a lot of new information. But the previous information is compelling, so the evidence is there.

Dale Bratzler: Any other discussion of evidence for the measure?

Man: This is (unintelligible) and (Taylor). Is there good evidence that early intervention gives you a different outcome than not intervening early?

Dr. (Scott Perek): Yes, very strong evidence that states that there's been some NIH-funded national trials that talked about early glaucoma intervention, and very obviously demonstrating that catching this early, lowering the pressure, dramatically changes the course of the disease.

Man: Thank you.

Dale Bratzler: Any other comments?

(Scott): As it correlates, there's also some - I don't treat glaucoma for a living, but there's also some evidence for also they've done some type of ocular hypertension, which are glaucoma suspects, and treating them even allegedly before they have full-blown glaucoma. It may be beneficial as well.

Dale Bratzler: Any other comments about evidence? So, (Asaba), do you want to go ahead and open up the poll?

(Asaba Denwulquafor): So, poll is now active on evidence. Committee members have the options to vote high, moderate, low, or insufficient. This is on measure 0086. We're waiting on three more votes. We're waiting on one more vote. The categories are high, moderate, low, or insufficient.

All committee members voted high or moderate on this criteria, so it's going to pass. I'll repeat the numbers shortly.

Dale Bratzler: Okay. So, gap?

(Scott): So, this measure has been used in public reporting, federal reporting since peak RS 2013, and although the performance rates very high and has gone down in 2017, from 95.4 percent to 90 percent, so there is some area for improvement for this measure, the performance rate, 2013 of 38.1 up to 2017, 85.06.

I, for example, don't report on this, because I don't do glaucoma. But most general ophthalmologists and glaucoma specialists report. In the AOA registry, the more registry, the 2017 performance rate was 53, up to 75 percent in 2018. So, it has gone up a little bit in their registry, but again, there's still room for improvement.

The previous discussion talked about the issues with African Americans and Medicaid population as well. The pre-evaluation comments were really not contributory. The only issue is the performance being above 90 percent. Is there enough of a gap to warrant continuation for reporting? Is there enough area for improvement?

Dale Bratzler: I'm curious, there's discrepancy between the various reporting modalities, QPP, registry, the old PKRS data rates were very high. I suspect that was selective or biased reporting by people that did well on the measure. Any rationale for these substantial differences?

(Elvia): Right, and I think there are - this is (Elvia). I think you're absolutely right, in that it has not - it is as voluntary measure, one it's choosing it and selecting is, those that would be - at least for QPP, those that would perform well select the measures.

I think that's why you see those high on performance rates. I think the registry, perhaps, is more indicative of the gap in care, because then there's more reporting to the registry.

Dale Bratzler: Any other comments or questions about gap? Then we'll go ahead and vote. Oh, I'm sorry.

(Kim): I don't want to delay or prolong our discussion. But I have glaucoma, open angle glaucoma, and I got a retinopathy expert, because I also have diabetic retinopathy, and macular edema, all the cool, fun stuff. And my doctor - my ophthalmologist treats retina, checks pressure.

To be completely honest, I'm a relatively with it patient in terms of - obviously I'm sitting in a room like this. I don't have the slightest idea to examine my optic nerve. How would a patient know this, if they weren't specifically told?

Dr. (Scott Perek): It's not clear, because it's something that you'd have to ask for other patients. But it's done. It's done routinely on every examination, and it's typically done especially in a patient with glaucoma, often, almost every visit.

(Kim): So, is that the kind of thing that, when he's looking at the back of my eye to see my retina, he would be able to see the optic nerve as well?

Dr. (Scott Perek): Correct. It's the same thing.

(Kim): Thank you.

(Scott): So, not to belabor the point, but even as non-glaucoma specialists, (Rishi) and I, we do a lot of diabetic retinopathy, and we routinely examine the optic nerve. I do it every single time on all the patients, and I document it every single time. If you look at my notes, it's in there, the evidence of the nerve was examined and the findings are all in there succinctly.

Dale Bratzler: Any other comments about gap? So, we'll go ahead and vote.

(Asaba Denwulquafor): Voting is now active for performance gap on measure 0086, primary open angle glaucoma optic nerve evaluation. The options are high, moderate, low, or insufficient.

Of the 18 committee members who have voted so far, four committee members selected high, and 14 selected moderate on performance gap for measure 0086, primary open angle glaucoma optic nerve evaluation. So, the measure passes on this criteria.

Dale Bratzler: All right, (Scott). Reliability?

(Scott): Reliability? So again, the theme for the afternoon is reliability was graded by the NQS staff as insufficient, for the same reasons. They looked at claims data

as well as the registry data, and the measure was found to be somewhat reliable.

But we have the same issues with the insufficient data for the reasons, again, that are - that were said prior, looking at the discussion about reliability. There wasn't a lot of discussion that wasn't already brought up, prior to that. I don't have any further comments at this time.

Dale Bratzler: I'm going to ask the developer, do you have any comments? I mean, it's the same issue, I know, that's come up on a number of the PCPI measures, that are PKRS-based or claims-based, so not having the individual clinician-level data.

(Scott): Yes, nothing else to add. Just pretty much the same issue with PKRS data. We're not able to determine what's at the group level versus the provider level, since we're getting both the tax ID number and also NPIs at the same time, so we can't discern between the two.

(Sam Tierney): This is (Sam Tierney) with the PCPI. Could I just add a little bit to (JR)'s answer? So, I think from - you know, I think there's two issues, potentially, here, you know, that are slightly different than the Hepatitis C measure.

There's this confusion over the one versus ten event, and you know, I've pulled up the testing results, and at one event, the reliability is .97, and at one event, the reliability is .98. So, if we ignore the ten-plus events, that's the reliability on the measure as it stands, and as it's specified.

The issue about the group or individual level, you know, I understand from a conversation we had with NQS staff, including Sam, a little bit ago, that because of us being unable to differentiate between the individual and group

clinicians within our data set that CMS provides to us, that we will need to have this considered at the group level, and that's - you know, that we can't necessarily have it considered at the individual level.

And I know for a measure we had recently reviewed by another committee, same exact similar issues with the data set, that was a decision that the committee made, and that was what was presented to the committee, is they - their vote was at the group level.

So, I guess I'd just ask Sam or other NQS staff to clarify if that same approach will be taken in this case, for us to ask the committee to vote at the group level, and then I - again, I just wanted to remind you of the reliability results of the one event.

Dale Bratzler: So, I believe the plan is to - as we vote on this particular measure, would be done at the group level. (Tim?)

(Tim): Hi, I just had a question about the first part of reliability, which data elements, if any, are not clearly defined. In part of the ICD10 codes that you have in there for primary open angle glaucoma, it includes glaucoma that's normal tension or low tension glaucoma.

And in some of the discussion about where this overlaps with other measures, I believe in another glaucoma measure, it really specifies primary open angle glaucoma, and I typically think of those patients as people having high pressures as part of their diagnostic definition.

Can you clarify why you included normal tension and low tension glaucoma in your overarching definition of primary open angle glaucoma?

Woman: Sure. I guess a little bit of history. I mean, these measures have been around for a very, very, very long time in PKRI, even when it first originated. And the value sets were developed in accordance with a variety of ophthalmologists, optometrists, and other providers from the expert work group.

I am not quite sure. I - from what I can see versus the other, I'd say, related measure, (unintelligible) measure that's in QPP, they also have the low tension in glaucoma. So, I think that that is consistent, especially because we have worked with those from AAO and AOA, to confirm the coding.

Outside of the other rationale, I know that it was mentioned earlier about potentially having the pre-glaucoma, but I don't believe that that is a focus of our measure here. I'm not sure that that provides a very strong answer for you, but it at least provides historical background and also considering the alignment with the coding for other measures.

(Tim): Thanks. I was just curious about the other ophthalmologists on the committee. Would you consider people to have normal or low tension glaucoma also to be primary open angle glaucoma, and if that is or isn't the case, why aren't we including everyone with glaucoma in this measure and just looking at people with open angle glaucoma?

(Scott): So the obvious answer is no. So, again, we don't treat glaucoma, but I'll have to go back and ask (Flora) when it was developed. I think this was a PDRP measure.

This is a super old measure, back when I first started doing this in the mid-2000s. So, I'm not sure why they were included. I'd have to go back and look

at it, but again, no, they're different disease processes and should be evaluated differently, measured differently.

(Sam Tierney): So, this is (Sam Tierney). I just wanted to comment, if I could, just to expand upon (Jaime)'s answer a little bit. I - you know, we do annual meet with our technical expert panel, based on feedback we get from implementation or evidence changes or any things of that nature.

So, you know, I think we've heard you on the coding, and I know we have reviewed the coding carefully with this expert panel, and maybe this issue just didn't rise to the top for them. But we could certainly review that with them carefully, as we go forward and - during our next round of annual updates.

Dale Bratzler: Okay, thank you. Other - any other comments about reliability? If not, we'll -

(William Tamper): This is (William Tamper) again. I just want to mention that the US preventive services task force found inadequate evidence that screening for or treatment of increased inter-ocular pressure or early detection of glaucoma reduces the number of persons who will develop impaired vision or quality of life.

Dale Bratzler: I think, though, wasn't that early screening versus somebody with an established diagnosis of open angle glaucoma?

(William Tamper): It's talking about screening, yes.

Dale Bratzler: Yes, so a little bit different.

(William Tamper): Are we talking about something different here, not screening?

Dale Bratzler: These are patients who have a diagnosis of open angle glaucoma, whether or not the optic nerve is evaluated at each visit, or, you know.

(William Tamper): Okay.

Dale Bratzler: Any other comments about reliability? Remember, we're voting on the measure, evaluating for endorsement at the group or practice level, because of the limitations of the analyses that PCPI is able to do on the data that they received from CMS.

(Asaba Denwulquafor): So, polling is now open on reliability for measure 0086, primary open angle glaucoma optic nerve evaluation. Committee members can select either moderate, high, low, or insufficient.

So, of all 18 committee members who have voted, two selected high, 15 selected moderate, and one person selected low on reliability for measure number 0086. The measure passes on this criteria.

Dale Bratzler: Alright, thank you. (Scott), do you want to go on with validity?

(Scott): So, validity is more interesting. So, the NQS staff gave it a grade of low. I think that may be the first one today. So, the validity was done using the typical testing, empirical validity, and face validity. And the face validity was, again, previously done, related from 2013 using 16 - an expert panel of 16 members.

Empirical validity was done looking at claims data and registry data. The registry data was higher at 0.57, and the claims data, there was a weak positive correlation of 0.22.

Face validity, this is again from 2013, the TEP agreed, that a strong agree - either agreed or strongly agreed the measurement adequately distinguished good and poor quality. So, there was adequate correlation with validity, and perhaps the NQS staff can elaborate on why they gave it a low grade instead of a moderate grade.

Man: Sure, thank you for the question. Now, the reason we gave it a low score was because of this correlation at the claim level, of 0.22. The traditional thresholds that we looked at were - we had the expectation of performing at a 0.3 level or above.

But that's up to the committee's discretion, as we understand that those sorts of thresholds aren't necessarily entirely set in people's minds for what those need to be, in order to establish that. And there isn't any guidance inside of NQS criterias as for what those correlations need to be, in order to qualify. So, we leave that up to the discretion of the committee.

(Scott): So, the question is, if you grade it at the registry level and you get a moderate score and the claims level, you get a low score, why do you grade the overall as a low score, not a moderate score, for example?

Man: Because the measure is specified for both. So, our concern is that, if it's - if we're just going to allow it for - only the registry would be entirely comfortable with that.

But if it were for - to include claims, well, then we have some concerns about validity of the claims validity data, in terms of its correlation with this particular external measure of quality that's proximate to the quality domain of interest.

Now, that's not to say that there aren't other measures that this might correlate strongly to, that - but that just wasn't analyzed by the developer.

(Scott): And again, it'd be nice to know how much is being reported using registry data versus claims data. If I was the majority as registry data, your point is moot. So, it should be a moderate or high.

Dale Bratzler: I think his point is there's two different ways that you could potentially report the data, and if you're doing it by claims, then as an NQS-endorsed major that has less validity, at least in the testing, than data that was submitted via the registry.

Man: And then the other question, was this data different than in 2013? Because presumably, it passed in 2013. Has it changed since then? Looking at the validity, has the - has it been reevaluated since 2013 and has it changed significantly?

(Scott): So, in 2013, we did not provide correlation analysis. So, this is new to this submission.

Man: So the 2013 submission just had the face validity.

(Sam Tierney): Yes, this is (Sam Tierney). I was just going to expand upon that. NQS requirements changed for validity, and so what was done in 2013 no longer meets their - our requirements of having empirical validity, which is why we did the correlation analysis.

(Nadine Chambers): And this is (Nadine Chambers) from the PCPI. And we had previously been told by the NQS staff that we should include both types of validity testing in updated submissions.

So, we provided both for the committee's benefit to see what we did in 2013, which, as Sam said, was face validity testing, and then we updated it to include empirical validity testing, which is what you see now under current testing data. Thank you.

Dale Bratzler: Any other comments about validity? Alright, we'll go ahead and open the polling.

(Asaba Denwulquafor): Polling is now active for measure number 0086 for validity. Committee members have the ability to select high, moderate, low, or insufficient. Waiting on one more vote, thank you. So, we have 18 votes in. 11 committee members selected moderate, and seven committee members selected low for validity.

Woman: That's consensus not reached, again.

(Asaba Denwulquafor): Which is 61 percent on moderate, and 39 percent on low.

Woman: Alright.

Dale Bratzler: Alright, feasibility.

(Scott): So, feasibility, again, this measure is using claims and registry data, and this is corollary measure 0086(e). And there's really not a lot of issues. The NQS staff gave it a moderate rating. There's no issues with feasibility for this measure.

Dale Bratzler: Any other comments about feasibility? Okay, we can open the poll.

(Asaba Denwulquafor): The poll is now open for feasibility. Committee members can select high, moderate, low, or insufficient. This is for measure number 0086, primary open angle glaucoma optic nerve evaluation. So far, we have 17 votes in, and all 17 committee members selected moderate.

(Sam Tierney): This is (Sam Tierney), with the PCPI. I'm so sorry to interrupt, but could I get some clarity on the last vote, on validity? I thought you said it was 61 percent, but you also said it was consensus not reached, and I thought the threshold was 60, just for -

(Scott): You are correct, and we did correct that in the room.

(Sam Tierney): Okay. Thanks, sorry. (Unintelligible) to see you, thanks.

Woman: Yes, wrong denominator on the math there, sorry about that.

(Sam Tierney): No problem, thank you.

(Nadine Chambers): And so, this is (Nadine Chambers). Just another point of clarification, then. So, does that actually mean that it passes validity?

(Asaba Denwulquafor): Yes, the measure passes on that criteria.

(Nadine Chambers): Thank you.

Dale Bratzler: So, let's move on to use.

(Scott): So, usability and use, first starting with use, the measure as you - we've already emphasized as being used for public reporting. It's used in our registry as well as the (unintelligible) registry, which is the American Academy of

Ophthalmology Registry, as well as the Moore Registry, which is the AAO registry. The NQS gave it a preliminary rating of S, and I don't have any further discussion for that.

Dale Bratzler: Any other comments about use? It is being used actively and has for many years, as I understand it, yes. Alright, we'll open the poll for use. We're - the afternoon phenomenon, I've seen this many, many times before.

(Asaba Denwulquafor): Polling is now open for use. We're voting on measure number 0086(e). Committee members can either select pass or no pass. So far, we've got 17 votes, and all 17 committee members selected pass for use on measure number 0086. The measure passes on this criteria, yes. That makes 18 committee members.

Dale Bratzler: I wonder who the 18th was. Alright, (Scott), usability).

(Scott): Again, to reiterate, the measure is currently being used in public reporting, in several registries. There's no issues with unintended consequences that anybody could discern, and the comments were non-contributory.

Dale Bratzler: The only question I had, not being an ophthalmologist, how subjective versus objective are these evaluations of the nerve head, the (unintelligible)? I remember learning about that in medical school, but how good is that?

(Scott): Yes, I don't - I don't know that. I think it's fairly objective, when you see a huge difference. But obviously, from my personal experiences, there's a lot of disparity. What I say is a small cup to dish ratio, my colleague would say is a large cup to dish ratio.

So, I think a better way of defining it would be take a picture of the nerve, which would be - you know, a lot of people do routinely, and then comparing apples to apples. But I mean, if there's a significant change, that's got to be pretty obvious, and I'm not familiar with the glaucoma literature to correlate that.

(Jaime): Just to be really clear, though, this is (Jaime). The measure does not actually ask you to look for specific responses, just that you have performed these exams and you have this sort of information available.

Dale Bratzler: I do understand that. I just - my whole question about this was does it make a big difference, you know, when it's - that was my kind of impression of it, at least just looking at the nerve head. I think the idea of photography makes a ton of sense. But how subjective is that process?

(Scott): That's the $64 billion question. I mean, we can get into that, and we can have future measures of carefully documenting progression. But before we actually document progression or look at the sensitivity and specificity of progression, you have to actually say you actually looked at the nerve. There has to be some evidence of that.

Man: That's what I was going to comment upon. I think if you look at where care may be shifting to more areas where perhaps optometry is providing care, places where there's not an ophthalmologist, a surgeon available, that the fact that this is part of a registry that is being measured now by optometry and that the rates were lower in the optometry registry than what's been provided in the ophthalmology registry or claims data for many years.

You know, maybe this is going to push people towards looking, and that's the first step towards addressing care disparity settings.

Dale Bratzler: Alright, any other comments? We'll open the poll on usability.

(Asaba Denwulquafor): Polling is now active for usability. We're voting on measure number 0086, primary open angle glaucoma optic nerve evaluation. Committee members have the ability to select high, moderate, low, or insufficient. So far, we have 17 votes in. Two committee members selected high, and 15 committee members selected moderate. The measure 0086 passes on usability. 18 now, 16 selected moderate.

Dale Bratzler: Alright, so we get to the final evaluation. Were there any public comments or related competing measures?

(Scott): So, there is one related measure, measure 0563, which is reduction of inter-ocular pressure by 15 percent. The documentation of plan of care, which is an outcome measure, although it deals with glaucoma, they're somewhat different.

Dale Bratzler: Alright. Any other comments?

(Scott): I think this is where my question kind of came up, about the eye pressure and what constitutes primary open angle glaucoma, because as it looks like it's written, both of them are addressing primary open angle glaucoma, but the eye pressure lowering measure, I think, only addresses those people that have high eye pressure to start, not necessarily a 15 to 20 percent reduction from someone that's already within the low range.

And at least as part of - as far as the comments from the NQS staff, you know, right here, that it seems that this particular measure includes other patients that have a different type of glaucoma than what's measured in that other measure,

and I'm - so I would just ask at some point for some clarification about why this particular definition was chosen for this measure, you know, and not the other measure.

And whether it makes sense then to just think about if you're going to look at the optic nerve as a factor in progression for glaucoma as a group of diseases, and there's many subsets of glaucoma, why do you choose to subset primary open angle glaucoma, right?

If that's one of your criterion in this case, and if you want to pick the same group as your eye pressure lowering group, then you should pick the same population. Don't include others, too.

Woman: Yes, thank you. And I think this is a measure - the IOP, 15 percent or greater reduction, is the measure I was referencing before, where we do have - they do include some of those low tension glaucoma codes as well in the denominator coding.

So, there is sort of that alignment across of those initial populations. I did just pull up the current QPC spec of that measure. It's (unintelligible) 141, and I don't see that there's a G code in there or something like that, that says you start off with a certain threshold of IOP.

So, I can't necessarily speak to that, but I - we definitely heard the comment, and we will be reviewing with the TAB, just to ensure that the coding is appropriate and this is something that we do want.

I think as Sam said on an annual basis, is review to make sure that we have the appropriate populations and appropriate coding capture, so that the measure

continues to do what it should be doing for the right population, right? So, I definitely appreciate the feedback and comments.

Dale Bratzler: Alright, thank you. So, any other comments? Otherwise, we'll open up the poll for overall suitability for endorsement.

(Scott): I have one quick comment. So, does 0563 include low tension glaucoma?

Woman: That is what I'm seeing on QPP 2019.

(Scott): Well, I mean, I don't remember. This has been around for, like, a decade, so it just seems logical that the title is incorrect. It's not primary open angle glaucoma. It is glaucoma.

Woman: But I think that you could argue that for both of them, then, for our measure and for AAO's measure.

(Scott): Right, so you may want to consider modifying the title of the measures to clarify.

Man: It just seems contradictory to call it a measure for primary open angle glaucoma and then include a number of diagnoses that ophthalmologists would not call primary open angle glaucoma. It would call it a different diagnosis.

Woman: Yes, and I will say that during the iterative process of updating this measure, perhaps at some point we included those codes and we started off as POAG, and just maybe just did not change the title of it. But that's something that we'll certainly look into, and as (Jaime) said, thanks for the feedback.

Dale Bratzler: Okay, we'll go ahead and open up the poll for suitability for endorsement.

(Asaba Denwulquafor): Polls are open for overall suitability for endorsement. This is measure number 0086, primary open angle glaucoma optic nerve evaluation. Committee members can either select a yes or a no for this measure. We're waiting on four more votes.

So far, we have 17 votes. Well, we have 18 votes in. 17 committee members selected a yes for overall suitability for endorsement, and one committee member voted no. Measure passes for this criteria.

Dale Bratzler: Alright. Thank you very much.

Woman: And just to backtrack, we never read the votes for evidence, even though we know it passed. Can we just verbally read them out, (Asaba), so that we capture them in the transcript?

(Asaba Denwulquafor): So for evidence on measure 0086, primary open angle glaucoma optic nerve evaluation, nine committee members voted high and eight committee members voted moderate, a total of 17 members. Measure passed on that criteria.

Dale Bratzler: Alright, thank you very much. I think we're turning it over to (Suzanne).

(Suzanne Tebar): Yes, for - we're going to now open the lines for public comment. We did receive one comment, a question via the chat box. Question for the measure owner, according to CMS published benchmarks, the average performance rate for the measure is 96, 97 percent for registry and claims reporting. This data is based on 2017 NIPS performance data. Where did 90 percent come from?

Woman: The 90 percent came from the 2017 CMS benchmark report, and so I'm not sure where the other numbers are being - have come from. But I can show you the benchmark, if you'd like.

(Suzanne Tebar): Sure. Maybe we can connect you with the chat questioner offline.

Woman: That'd be great.

(Suzanne Tebar): We'll follow up. Any other questions, either on the phone or in the room? Alright, (Asaba) will take us through next steps.

(Asaba Denwulquafor): So for the next steps for the primary care and chronic illness project at NQF, we have a committee post-measure evaluation web meeting. This is going to be a two hour meeting on Tuesday, July 1st, from 2:00 to 4:00 pm Eastern time.

We'll also have draft reports and the comments period, which will run from August 1st to August 30th of 2019, and the post-comment web meeting, which is right now, dated for September 24th, from 2:00 to 4:00 pm.

CPAC reviews for the current measure that have been evaluated is going to be in late October, early November, and the appeal period after CPAC is going to be from November 6th to December 5th of 2019.

The measures that have not been discussed during the meeting today will be discussed at the post-measure evaluation web meeting. I'll now pass it over to Adam, Dale, and Sam for the closing remarks.

(Sam Tierney): Actually, on Monday, July 1st. Apologies for that typo. We'll be talking with you Monday afternoon, not Tuesday.

Woman: Yes, and if anyone is not able to attend, please email us right away, because we do need quorum to vote on the measure. So, please reach out to our team if you know you have a conflict.

Man: Adam and Dale?

Adam Thompson: I just want to say thank you to everyone for sticking with us, all the great discussions and conversations. I know we didn't get quite finished today, but it was an ambitious agenda to begin with, so we look forward to speaking with you all again on Monday, and thank you for all the hard work as well as the measure developers who came in person and on the phone.

Dale Bratzler: And I have nothing else to add to Adam's comment. Thank you all.

Adam Thompson: And speaking to the NQS staff, a big thanks to all of you for all the hard work, especially the measure developers for grace under fire, and for what's undoubtedly a very heavy lift in putting together these submissions. It's more appreciated than you probably realize, and thanks for working with us. I wish you all safe travels home, and we look forward to talking to you on Monday. Thanks very much.

Woman: Meeting is now adjourned.

END