NATIONAL QUALITY FORUM

Moderator: Benita Kornegay Henry
June 27, 2019
11:00 am ET

(Ashley Wilvan): Good morning everyone. It looks like we're getting close to a critical mass here. And so I think we will go ahead and get started and see where we are with attendance so far. Again, this is (Ashley Wilvan) from NQF. I just want to welcome everyone for joining us today. We have two webinars scheduled today to evaluate three cost measures that have been submitted by CMS and the Acumen team. And we'll do our best to try to get through those three measures today. And we'll - under the - with the guidance of (Cheryl), our co-chair today and we will, again, like I said, do our best to get through those measures today. If in fact we don't, we do have another call scheduled as back up in July, but we'll do our best to keep things moving.

A couple of housekeeping items to get us started. We are using the webinars today, so feel free to use the raise your hand and the chat feature to communicate when you'd like to speak or if you have comments you'd like to be considered. We will be watching that chat box as well as the hand-raising features to make sure that folks are called on in a timely manner and that everyone can have an opportunity to speak.

We will also ask that you mute your line when you're not speaking. Again, it helps to keep down on background noise and any interference and feedback that we may get with a lot of open lines. So I did want to welcome one of our new team members. (Kate Buchanan), who's joined us, (Kate), do you want to say hi?

(Kate Buchanan): Sure. Thank you so much (Ashley). My name is (Kate Buchanan). I'm a senior project manager here at NQF. I've been here for about three years working primarily with the measure applications partnership, and I'm really excited to engage with the standing committee on cost measures.

(Ashley Wilvan): Thanks (Kate). So we're happy to have (Kate) on board. Of course I've worked with the cost project for some time. (Navea Kumar) is our project analyst who's been working with us for some time as well. And I think you guys are all familiar with her via email at this point. (Saroun Amin) also continues to consult and help us out on this project. (Saroun), are you there?

(Saroun Amin): Yes, (Ashley).

(Ashley Wilvan): Hi, (Saroun), do you want to say hi?

(Saroun Amin): Oh, hello everyone. Excited for another round of measure evaluation with the standing committee.

(Ashley Wilvan): Great. Thanks (Saroun). So for today we'll again - our conversation will be focused on consideration of the three candidate measures. We do have a little bit of - kind of portfolio discussion at the end. Hopefully we'll get to that across the - we'll have time for that at the end - hopefully the end of this afternoon's webinar. If not, we'll find another time to discuss it and get the committee's input.

But, we'll save that for the last with the focus being on getting through the three measures today for evaluation. So, let's go ahead and roll into the roll call. Just a brief note with the committee. You will notice that we have a few, some few, fewer people, so that we have had some folks that have had to sit down due to competing priorities. We will have an upcoming nominations period in the fall to fill some of the seats that we've lost. But for today, this is the committee that will be working with us through this cycle.

And we will be voting and do our evaluation with the group that we have listed here. So I do know that (Brent) had some traveling challenges and will not be able to hop on the phone this morning. So we will miss him. We will also be incorporating our disclosure of interests into the roll call today. So, I'll do a brief overview of our disclosure and then, as I call your name, if you could just let us know if you have any interests to disclose. And then we'll go from there.

So, you recall that we distributed this disclosure of interest form when you first came to the committee and then we send you one every year to update, after that. And then also for each cycle we send out a measure-specific disclosure interest form that asks you about your relationship to any measures under review and then any other related or competing measures that we identified.

So across those two forms, we asked you a lot of questions about your professional activities and the degree of your involvement with any of the measures under review. So today, like we do on all of our measure evaluation calls, in the interest of transparency we're going to ask that you orally disclose any information that you provided in either of those forms that you believe is

relevant to this committee and specifically the matters under review, for the - any related or competing measures, which I do not believe there were any.

Again, we don't need you to summarize your resume. We're just especially interested in any work that is specifically related to the measures under review. That would include grants, research, consulting or measure development activities.

And so just a couple of reminders. You sit on this group as an individual and do not represent the interests of your employer or anyone that may have nominated you. Again, we're interested in disclosures of both paid and unpaid activity that is relevant to today's work. And just because you disclose does not mean you have a conflict of interest. Again, we do just oral disclosures in the spirit of openness and transparency.

So starting with (Cheryl), I'll call your name and let us know if you're here and if you have any interests to disclose. (Cheryl)?

(Cheryl Sandberg): Yes. Yes, can you hear me?

(Ashley Wilvan): We can.

(Cheryl Sandberg): Great. I am here and I'm happy to be joining the call today. I do not have any interests to disclose.

(Ashley Wilvan): Thanks (Cheryl). (Kristin Martin Anderson).

(Kristin Martin Anderson): Hi. I am here am and I do not have any interest to disclose for these measures.

(Ashley Wilvan): Thank you (Kristin). (Troy)?

(Troy Theisinger): This is (Troy Theisinger), no interests to disclose. And I'm driving between meetings. I won't be able to log into the webinar for about an hour.

(Ashley Wilvan): Okay. Thanks for the heads up. (Nancy Garrett)? Okay. (Andrea Gelder). (Rachel Howe)? (Sonny Gemnani).

(Sonny Gemnani): Morning. This is (Sonny Gemnani), I have no disclosures to - of interest.

(Ashley Wilvan): Thank you. (Lisa Last). (Lisa), are you there? (Jason Lot)? (Jack Needleman).

(Jack Needleman): Good morning. Nothing close.

(Ashley Wilvan): Thanks (Jack). (Janice Ulowski)?

(Janice Ulowski): Present, no disclosures.

(Ashley Wilvan): Thank you (Janice). (John Ratliff)?

(John Ratliff): Good morning everybody. Present, nothing to disclose.

(Ashley Wilvan): Thank you. (Srinivas) just sent us a note. He had a last minute conflict so he won't be able to join us this morning, will hopefully join us this afternoon. (Lena Walker).

(Lena Walker): I'm here and I have no interest to disclose.

(Ashley Wilvan): Thanks (Lena). (Bill Weintraub)? Okay. (Herb Wong)?

(Herb Wong): Yes, I'm present and I have nothing to disclose.

(Ashley Wilvan): Thanks. Welcome everyone. So with that…

(Lisa Last): So this is (Lisa Last). I joined and I have nothing to disclose.

(Ashley Wilvan): Hi (Lisa). Thanks, welcome.

(Lisa Last): Sure, hello.

(Ashley Wilvan): So, we'll go ahead and before I hand it over to (Navea) to give an overview of the voting and some of the procedures for today, I did want to obviously welcome (Cheryl) as our co-chair today and see if you have any opening remarks for the group before we jump in.

(Cheryl Sandberg): Thanks (Ashley). I just want to say thank you to all the committee members for attending today's marathon back-to-back meetings. I know this is a big time commitment but it's very important to the larger departments' measurement enterprise. So, despite the fact we're all contributing a lot of our time, I think this is a critical endeavor and your input is greatly appreciated.

We're going to try to move efficiently today because I know people have, you know, other things that are high on their priority list to get to. But I also want you to thank the Acumen team for all their good work and I'm glad they can join us today to help, you know, clarify any questions the committee may have. Back to you. (Ashley).

(Ashley Wilvan): Thanks for that (Cheryl). And I would also like to check in with the Acumen team to make sure you're there and that we can hear you. There'll be giving us an overview at several points in the call. Are you there?

Sri Nagavarapu: Hi, (Ashley), this is Sri for Acumen, we're here. Can you hear us okay?

(Ashley Wilvan): Yes, we can. Okay, thanks. We'll check in with you with you shortly to get started on your overview. For now I'm going to hand it over to (Navea) to talk through - (Navea) and (Kate) to talk through some of the operational and logistical parts of our call today.

(Navea Kumar): Hi. Thank you (Ashley). This is (Navea Kumar) speaking. So today I'll just start with getting the voting platform information. We have sent over an email yesterday afternoon with the instructions for the poll everywhere platform. You should be having it in your mail with the link for the Web site. So we will be conducting voting today, I believe, and that could be accessed only by your computer. You should be able to see a welcome screen with the poll as will appear once the (unintelligible) opens voting.

After you place your vote, you'll see a message stating that your response has been recorded on the screen. Please let us know either through the chat box or email if you're having any difficulty once we begin voting. Is there any question related to voting?

(Ashley Wilvan): Did everyone get that email or, have you been able to locate that email? I know that (Troy) is driving so he probably won't be able to do that now. But are there - for those of you who are in front of your computer, could you give us a sense of your ability to access that?

Woman: No problem for me.

(Lisa Last): This is (Lisa), I got it.

Man: No problems here.

(Jack Needleman): This is (Jack), on a different place than I usually am. And I'm on a guest computer here and I think I'm logged in, but I am not seeing anything. I'm seeing a gray screen. Is that what I should be seeing?

Woman 2: I see a blue Welcome to NQF vote presentation.

(Jack Needleman): No? Okay, I've got to go log in again. I will go do that. Thank you.

(Lena Walker): I'm sorry, what am I supposed to be seeing on the screen? This is Lena. I see the slide that says voting prep. Am I supposed to see a different slide or different screen?

(Navea Kumar): It should be a welcome screen for NQF vote two, T-W-O.

(Lena Walker): No, I don't see it. I just have slide deck on my screen.

(Navea Kumar): We can resend you the email Lena if you …

(Lena Walker): Thank you.

(Navea Kumar): Okay. So, while (Kate) is helping out Lena, I'll just continue with the next slide. So measures review. This is currently our portfolio review. We - the measure of 3474 has been endorsed by feedback as of June the fifth, but hospital-level standardized payment associated with a 90-day episode of care for elective primary total hip and/or total knee arthroplasty.

And so today we'll be reviewing three submitted measures. They're new measures. They were all developed by BMS and Acumen. First is measure

3590, (unintelligible) cataract removal with intra-ocular lens implantations. Measured 3510, screening/surveillance colonoscopy, and measuring 3512, knee arthroplasty.

And later on, during this meeting, developers like (Ashley) said, will have some time to provide an introduction to each measure along with an overview of their methodology and clinical logic.

And with that I will pass it to (Kate) to go over the measure evaluation process as well as the voting process.

(Kate Buchanan): Great. Thank you. And so I want to take a little bit of time to talk about the measure evaluation process. So if we move forward one slide. So one of the things is that there are - as you know the cost and efficiency committee is one of 15 topical multi-stakeholder committees. The standing committee considers feedback from three different sources identified on the slide.

The first is the scientific methods panel, which consists of statistical and methodological experts. The scientific method panel reviews the scientific acceptability criteria.

The technical expert panels are another source of input the standing committee considers. The CUPS consists of clinical expertise and evaluates the clinical elements of the measure. The final source of which the - input that the standing committee receives are public comments and member support that provide a multi-stakeholder perspective.

Also on the next slide. You can see for the spring 2019 measure evaluation. All measures are first reviewed by the scientific methods panel. The scientific methods panel rated the measures as not passed, consensus not reached, or

passed on the scientific acceptability criteria. Regardless of the rating the scientific methods panel gave, the clinical test reviewed and provided feedback. For the measures that did not pass the scientific acceptability criteria, the feedback provided by the scientific methods and the tests went to the developer.

For measures that the scientific methods panel rated as consensus not reached or passed, it'll be reviewed today by the standing committee.

The scientific methods panel is comprised of individuals with statistical and methodological expertise. The panel was established to ensure consistent evaluation of the scientific acceptability of complex measures. It evaluates the reliability and the (unintelligible) of the measure. This cycle, eight measures went to the scientific methods panel. Five did not pass due to issues related to the reliability testing. For the three measures at the scientific methods panel passed on reliability, the panel did not come to consensus on validity.

(Ashley Wilvan): Hi (Kate), this is (Ashley). Just before we move forward to that next slide, I just wanted to point out a couple things. We are in the process here at NQF at looking at the process and the transition of measures between the methods panel and the standing committee to make sure that we're being transparent with all the information that the methods panel produced in terms of their evaluation. And so I did, you know, going forward we'll be sending a lot more information forward to standing committees in terms of the methods panel reviews that summarizes their discussion and really some of the key issues that came up.

We are lucky enough to have (Jack Needleman) on our committee, who is also a member of the methods panel. So, what I'm going to do, sorry to put you on this project, what I'm going to do is give us a really high level overview of

some of the things that - some of the issues that came up with the methods panels review with the five measures that did not pass. And I did also want to clarify that the consensus not reached issue on validity actually carried forward for all eight measures. I think the - with maybe with the exception of one that did not pass validity as well, but essentially I think the methods panel really struggled in general with the empirical validity approach.

But for the five measures that didn't pass due to reliability, as you can see with the three measures that you reviewed, the methodology for detecting was very similar across all eight measures. They did a similar testing approach in terms of the test/retest with correlation, the signal to noise, and then an assessment of the quintiles and movement or rankings across quintiles.

So in that assessment, I think one of the key issues that the methods panel struggled with was the actual scores with the signal-to-noise and some of the correlation scores that the methods panel has some discomfort with in terms of the - of how low some of them were based on the sample size that was presented in the forum.

And so certainly the developers have gotten feedback about that and we are hoping that they will resubmit those in a future cycle. So I just wanted to kind of provide a brief overview of that, and (Jack) I don't know if you have any other details you'd like to share with the committee on maybe some of the concerns with the measures that did not come through.

(Jack Needleman): No, I think the ongoing issue across a lot of the measures was the level of reliability, either measured signal-to-noise or some other measure that would meet the standard for endorsement. There was a lot of discussion in the individual subgroups that were working on individual measures. And this was a major source of discussion at the full committee meeting held last week or

the week before. So that came up. I think the issue of substantive validity where clinical expertise needs to come in was often kicked down the road to this committee.

(Ashley Wilvan): Thanks, (Jeff). So with that, I just wanted to see if there's any questions from other folks on the phone about - specifically about the measures that did not pass and otherwise we can keep rolling. But I just want to give an opportunity for the committee to respond if there's any questions.

(Lisa Last): This is (Lisa), for the measures that did not pass, are they measures that are being used already by someone somewhere?

(Ashley Wilvan): I may ask the Acumen team to respond, but I believe all of them are slated for myth. But I'm - I would probably have them explain exactly where they are in terms of their use within that program. So I don't know, Sri, are you able to address that question?

Sri Nagavarapu: Yep. So all eight measures, the three that you'll see today as well as the other five went through the notice and comment rulemaking process for the 2019 performance period, this performance period last year. Went through the public comment period and were finalized for 2019 performance group.

(Ashley Wilvan): Any other questions about those measures? Okay, we'll go ahead and keep moving.

(Kate Buchanan): Sounds great. So here we have the seven clinical cuts that reviewed the eight measures this cycle. The cap is charged with review and providing feedback on clinical logic, episode trigger and end definition, clinical exclusion, clinical risk factors for risk adjustment. This is qualitative feedback only and feedback is in the PAs for the three measures discussed today. And for those measures

that did not pass the scientific accessibility standards, staff will summarize and send the developer.

So for our evaluation for today. Here you can see kind of our ground rules for today. Keeping comments concise, trying not to repeat what other people have already said. That's, you know, obviously basing our evaluation recommendations on the criteria and guidance.

The process for measure discussion. So the - there'll be a brief introduction by the measure developer, about three to five minutes. Then the lead discussants will begin committee discussion for each criteria. Because there was a lot of similarities in the way that the measure measures were developed, we have the same people evaluating all three of the criteria for each of the - all of the same criteria for each of the three measures.

So the people, the lead discussants for reliability are the lead discussants throughout all three use and usability, et cetera. So our lead discussants will briefly explain information on the criterion provided by the developer. Provide a brief summary of the pre-meeting evaluation comments from the tech, scientific methods panel or other committee members. Emphasize any areas of concern or differences of opinion, noting if needed the preliminary rating by NQF. And this rating is intended to be used as a guide to facilitate the committee's discussion and evaluation.

The developers are on the call and are available to respond to questions at the discretion of the committee. The full committee will discuss, then vote on the criterion if needed before moving on to the next criterion.

And so as I said, we have lead discussants who are assigned to each criterion. They will begin the discussion of the measure evaluation criterion including

summarizing the evaluation of each criterion and based on all of the standing committees' pre-meeting evaluation comments, highlighting areas of concern or difference of opinion. They will verbalize conclusions regarding how well the measure meets NQF evaluation criteria, be fully conversant with the submitted measure information on the assigned measure criteria, and then for the discussants we ask that they be fully conversant with the submitted measure information on their assigned criteria and supplement the lead discussant comments with evaluation remarks as needed.

(Ashley Wilvan): So thanks (Kate). (Kate), this is (Ashley). Sorry, before we move on, I did just want to point out that for some of the folks that were not able to join us today, we may need to do a little bit of shuffling for lead discussants. So, for important, Lena, you were the lead and I don't believe that I've heard Andrea there. (Jennifer) has resigned since we sent this out initially, and I don't believe I heard (Rachel) unless she's joined since we started.

So, we'll have Lena lead things off and then we'll kind of jump into the discussion after that. For reliability, (Srinavas) was the lead discussant and he had a last-minute conflict. So we do have (Herb) and (Lisa Last) on the phone, so hopefully - I know you guys weren't assigned as a lead, but if you guys feel comfortable at all with giving a brief overview in (Srinavas)'s place, that would be great.

Otherwise, between (Cheryl) and I, we can try to jump things off a little bit. Validity, I believe we have (Sonny), (Jack) and (Troy), so we should be good there. And then we should be good with feasibility and usability and use as well. So I just wanted to just give a quick overview, particularly for importance and reliability where we've had some absences that may impact the discussion.

(Kate Buchanan): Great. Thank you so much (Ashley), I appreciate you doing that. Well, if there are no questions about the lead discussants, we can move on quickly to review the voting process. So here you can see the endorsement criteria. The importance to measure report, which is a must pass. That's the opportunity for improvement in vote, we have the scientific acceptability. Also a must pass criteria.

For the reliability, the scientific methods panel did not - voted on the reliability for all three of the measures and came to conclusions. The committee can either choose to accept the scientific methods panel vote or they can revote. It's at their discretion. For validity, the committee must discuss and vote on validity since the scientific methods panel did not reach consensus.

We'll then move on to feasibility, usability and use, and then overall suitability for endorsement. If any of the measures fail on one of the must pass criteria, there is no further discussion or voting on the subsequent criteria for that measure. We will move on to the next measure.

And so, I apologize, we actually have an updated quorum. So our quorum is ten people and that will help us reach 66% agreement. And right now we have ten committee members on the call, but I know that several committee members are not able to log in to the polling at this moment. So I wasn't sure. (Ashley), I know we had talked about a couple of different ways to proceed, but I wasn't sure exactly how we wanted to go forth with the voting.

(Ashley Wilvan): Right. So, I think the - maybe the easiest way so - would be to probably use our backup until (unintelligible) because (Troy) is our tenth person. He's not able to get to a computer right now. We need him to have a quorum. What we will do is send out a Survey Monkey link that you can just click on the link in

the email and you can vote on your own computer. And we will collect those votes after the call.

The one tricky part we may need to navigate a bit is obviously if something fails, we won't know right away. We'll do our best to try to track that real time. But until we have forum on the call, we won't be able to use the voting poll software that we discussed earlier because it actually shows on the webinar screen what the kind of total votes are. And so, we don't generally use that if we don't have quorum.

So we'll update you as we go. But for now, let's just plan to use the Survey Monkey link that (Navea) will send out to you shortly. It'll be at the top of your email and we'll just have everyone vote along as we go through the discussion.

(Kate Buchanan): Thank you so much (Ashley). And the Survey Monkey link was just sent out. So it should be arriving at the top of everyone's inbox. It has a high importance on it. And if you click on the link for survey, it'll take you directly to the Survey Monkey link.

So then, I guess we can skip the voting test. We can go back to that if we reach quorum on the call. Does that sound okay, (Ashley)?

(Kristin Martin Anderson): Hey, this is (Kristin). I just wanted you to know, I haven't received any new emails, so maybe it's still coming. I have the original survey email from last - wait let me see.

Woman: I just got it like 30 seconds ago.

(Crosstalk)

(Kristin Martin Anderson): Does that mean we don't - we ignore the other (unintelligible) Web site?

(Kate Buchanan): Yes. For now, we will ignore the other one.

(Kristin Martin Anderson): Okay, got it.

(Ashley Wilvan): Yeah. Maybe just minimize it for now, but we'll use Survey Monkey. Yeah, thanks. Sorry for the confusion. Sometimes it's a game time - depending on - game time decision, depending on where folks are on the call and (unintelligible).

(Kristin Martin Anderson): Okay. Thanks. Got it.

(Ashley Wilvan): …came through for you, (Kristin)?

(Kristin Martin Anderson): Yes.

(Ashley Wilvan): Okay, perfect. Thanks.

(Kate Buchanan): Great. So for right now, we can skip the voting test, and (Ashley) should we move forward to the consideration of candidate measure?

(Ashley Wilvan): Yes, let's do that. So (Kate), we'll start with the measure developers giving us a high-level overview of their measure constructs and measurement approach. We did have an orientation call back in May and I know only a few of you guys were able to make that. And so you probably missed a good portion of that. We do think it's a good foundation for the discussion. So we did, have Acumen give kind of a condensed version of that overview for today before

they move into a brief discussion of the clinical logic, specifically for the first measure, which will be 3509.

So, they'll give about a, I would say, you know, 10- to 15-minute overview starting with the overall measure construct for that - applies to all three measures. And then they'll give an overview for 3509. We'll discuss 3509 in its totality and then they'll give a brief overview for the clinical logic for 3510 before we get started and the same for 3512. So, with that I will just hand it over to the Acumen team to get us started.

Dr. Choradia: Great. Thanks (Ashley). So hi everyone. Thank you for joining us. We really appreciate the opportunity to speak about these measures. On this call we've got a number of team members from Acumen that are going to be helping to clarify any questions about the members. So I wanted to introduce our - the measures. So I wanted to introduce everyone.

My name is Dr. Nirmal Choradia, I'm an internal medicine physician. We also have Dr. (Suzanne Pershing), an ophthalmologist and Dr. (Rose Doe), a cardiologist, and Dr. Sri Nagavarapu who you've already heard from, who is the statistical adviser and project director.

We really want to illustrate the overarching process for all the measures and later on prior to the discussion of each measure we'll go into the specifics. So first I'll talk a little bit about the process.

The process for measure development for the three episode-based cost measures being discussed today consisted of an eight month intense effort across three expert panels which comprise 70 clinician subject matter experts affiliated with 44 different clinical societies.

These experts were - are distinct from the clinical tap that the NQF convened and had extensive time to review the measures thoroughly, consult with their societies and come to consensus recommendations regarding the construction of these measures.

We convened these expert panels based on clinical areas, and across the three committees we had cataract surgeons, GI doctors, general surgeons, colorectal surgeons, orthopedists, physical therapists, geriatricians, and many others that contributed their specific expertise that was applicable.

And so during this eight month period, we met with each panel of clinical experts across six to eight multi-hour meetings via webinar as well as in person in order to select episode groups that were felt to be high impact and clinically important, define the timeframe for the episodes and delineate all of the elements to be included in the cost measures. We'll also create subgroups and define risk adjustment.

The specifications were systematically tested for scientific accessibility at each stage of measure development with the clinical experts. And we provided them with analyses to help inform the decision-making process on all of the specifications.

All of these findings were discussed extensively during these meetings and measure specifications were implemented when consensus was established with greater than a 60% consensus of subject matter expert committee members voting in favor to mirror NQF voting practices.

And so beyond the committee, the measures underwent a national field test for all clinicians who would be impacted by the measures. This was something as far as we know that is unprecedented as far as field testing. We shared the

feedback reports with every clinician and clinician group that was attributed at least ten episodes. And for these three measures, between 3,000 and 6,700 group practices and 7,600 and 19,000 individual clinicians received reports.

We can't give - we don't have a unique number across these three measures, but the upper bound you can think of has 14,000 clinician groups and 37,000 individual clinicians.

So we received public comments during the month of field testing and brought them back to our subject matter expert committees to consider all potential replies. These measures were proposed and finalized for use in 2019 rulemaking and the episode-based measures have been really highly in demand by stakeholders. During that time they went through an additional two-month public comment period. And the measures are weighted along with other measures in the MIPS cost categories in order to construct the MIPS final score.

They provide an important complement to the population-based all cost measures which are currently used, TPCC and MSPB. And each measure will also undergo annual maintenance and will undergo comprehensive reevaluation on the annual cycle.

So I had a couple of points of emphasis that I really want to get to. We reviewed each measure worksheet, and we feel that the points that you all have raised are amongst those that have also been discussed in detailed by our committees over the eight-month development process.

In addition, we've provided data analyses to help inform those discussions. And so the decisions that the committees came to were not made lightly in these respects and acknowledged all of the - many of the possible tradeoffs.

We're definitely happy to highlight the discussion points on any of these topics if you would like us to like to go into more detail.

And what we're really discussing today is a measure for clinicians built by clinicians after months of significant deliberation, and nationwide testing and comment. And another point that we want to make about risk adjustment. Not only do these episodes have an extensive number of risk adjustment variables based off of the CMS HCC model, but they also include risk adjustment variables that the committee decided was important and specific to each of the episodes, such as including a risk adjustment for traumatic arthritis and knee arthroplasty.

In addition, the committees were also able to subgroup and exclude patient populations in order to make a more homogenous population for comparison. And the last piece of risk adjustment that isn't really discussed as a risk adjustment but actually is, is the ability to only include services related to the specific procedures. This is a risk adjustment methodology that has not really been previously utilized, but what it does is that it removes variability that's outside of the clinician's control.

So as you all have probably realized, you know, in order to ensure consistency and transparency, we had a similar framework across these measures, but each measure is clinically distinct, similarly to how each of the procedures are clinically distinct. And really quickly, I wanted to go into an example of the detailed discussion surrounding certain points. And specifically I wanted to go over the GI committee's discussion of whether or not to include pathology.

On the one hand, against inclusion of pathology costs they discussed how inclusion of pathology could inordinately affect clinicians who identified more precancerous lesions and have a higher adenoma detection rate. And

while they did acknowledge how this concern was mitigated slightly given the considerations they gave to exclude certain populations that have a much higher risk of polyp development, in order to kind of get at that higher polyp rate case mix concern, this was definitely one of the concerns that they expressed.

On the flip side, the committee was concerned about known practices of spreading biopsy specimens across multiple pathology bottles versus a single bottle, which would greatly increase pathology costs without improving - without really improving identification of cancerous lesions. Also, they were concerned that colonoscis may be over biopsying and removing noncancerous polyps and sending them off for pathology.

This concern was augmented by the fact that there are newer GI discussions surrounding removal and tossing of identified noncancerous polyps. In conjunction with these discussions, Acumen provided analyses of pathology cost occurrences within much larger analysis in order to provide them data to help inform this decision.

And after considering the trade-offs and taking into account - and also taking into account the adenoma detection rate quality measure that will hopefully be in MIPS, the committee voted with greater than 60% consensus, specifically 64% agreeing to include pathology costs because they believe that this specification along with other aspects of the measure they provided input on would help the measure differentiate between high and low performing providers.

So in conclusion, these measures for clinicians have been developed by their own peers in a well-defined, transparent process. The experts we work with realize no measure is perfect and that every measure has tradeoffs. Their goal

was to create an actionable cost measure that could be used for themselves and their peers in MIPS in combination with quality measure. These measures are the result of their long effort to ensure the relevance of these measures within the MIPS program.

So with that, that's my brief introduction on these and I wanted to go ahead and pass it back to you all. And again, the Acumen team, all of us here will be happy to provide any clarifications or any further insight into the discussions that were had during the committee meeting.

(Ashley Wilvan): Well, thanks Sri this is (Ashley) from NQF. I wondered, before we have the Acumen team give a brief overview of the cataract removal measure, if there's any questions about the measurement approach that Sri discussed that applies to all the measures of kind of the underlying measure construct that you have any questions about before we dive into the individual matters.

Lisa: Yeah, this is Lisa, I have a question. Are the specialty societies that participated in the process, are they endorsing these measures?

Sri Nagavarapu: This is Sri from Acumen. The specialty societies voted on each stage of the measure specifications as the measures were developed using the 60% threshold. So that’s sort of a case where they've kind of seen each aspect of the measure and indicated approval. And there are going to be cases where there's tough decisions to be made or where maybe 70% of the committee agrees on something and 30% disagrees. And that there’s definitely situations that like that, that came up.

But we felt it was important to have a 50% threshold on sort of the decision to ensure that the societies felt that these were measures that provided actionable information and opportunities for improvement. And the episode-based

measures have been highly in demand by especially societies and stakeholders.

And this was a vehicle that we felt was the way to get those societies to the table and make these sorts of very difficult decisions that the reviewers on the NQF panel as well as the methods panel and NQF body have brought up, because I think that's really the only way to, to make those types of tough decisions is with the people that are actually affected by them and the ones that are demanding these episode-based measures.

(Crosstalk)

(Cheryl Sandberg): This is (Cheryl). Can I jump in and ask a question?

(Ashley Wilvan): Sure, please, go ahead.

(Cheryl Sandberg): … across all three measures? I'm curious for the Acumen team. As I was reading through all the materials for the meeting, it seemed that when each of the measures was reviewed by the TAB, the clinical members, that they raised what I consider to be significant issues around certain things that were included that potentially should be excluded.

And, it seemed as though, and I just want to make sure my understanding is correct, that some of those items were sort of kicked down the road to sort of see kind of after implementation whether or not, you know, they might be revisited. Could you clarify how you handled the different clinical issues that were raised by the top members?

Sri Nagavarapu: Sure. I would be glad to, this is Sri again from Acumen. First off, I just want to make an important distinction. The clinical tap comment that you all saw,

the comments from the tap that NQF convened, that tap had a chance with a small number of members, the handful of members were able to review the clinical specifications. I think they had a week or two to review the clinical specifications.

That tap is distinct from the eight month-long process that Nirmal just talked. And so the tap that we have for the tossed measure work as a whole to provide strategic vision for it as well as, the body of clinical experts that were convened in these clinical subcommittees that do that measure development and that we're fielded from nominees from the specialty societies, that was a distinct group from the clinical tap comments that you saw.

So the types of comments that you saw for the clinical tap, those were comments that I think are very important. But at the same time, as Nirmal mentioned during that eight-month long process that came before them, those comments had been considered. And it got through the clinical subcommittee process from the specialty societies and decisions and votes were made on those types of issues.

And so, you know, while we do have an annual measure updating process and there are consistent opportunities to revisit decisions, I do want to make that important distinction between the NQF-convened clinical test versus this kind of eight-month-long process that Nirmal was talking about. That's a separate group of experts that was actually constituted from the specialty society nominees themselves.

(Janice Orlowski): So, this is (Janice Orlowski). I've been listening carefully to what you are addressing with the last two questions. But I have to say that it's not clear to me. So I do understand that there has been a vote of greater than 60% by the TAP that, you know, brought this forward. But I believe the first question was

do the specialty societies that you worked with - have they endorsed these three measures? And this - so the question in my mind is kind of a yes or no thing. And maybe yes for some and no for others. But I'm not sure that I heard the answer clearly.

Sri Nagavarapu: Sure. We did not go to each specialty society itself and ask for a vote. What we did was look at the votes from the committee that was constituted by individuals, including those nominated or recruited from the specialty society. So it was the committee members who were affiliated with the specialty societies that voted and had the opportunity to discuss the issues with their specialty societies during the course of development.

For instance, with GI, with the colonoscopy measure, there was very close involvement on the field testing period between the clinical subcommittee members and their societies to make updates to the measures that were incorporated into the metrics. So, but it is the committee members that voted on this, not the society themselves.

(Ashley Wilvan): Thanks. Thanks for that. Are there any other questions before we start with the first measure?

(Jack Needleman): Yeah, this is (Jack Needleman). And first thing I want to say Sri is, while I may be critical of some of the measures during the meeting, this is not a criticism of you and your team. I think you've done extraordinary work, looking at the data you had to make the decisions you made in analyzing the measures. So, CMS has been extraordinarily well served by you and your team.

We've been talking about the consensus process and I just want to understand what the committee had before them with respect to reliability. For example,

did the committee see the same signal-to-noise analysis and test, retest and all the other stuff that is in the NQF packages as they were thinking about the reliability issue? And similarly on the validity issue, did they have - did they see the same analysis of correlation of cost components with the measure that is presented as part of the validity analysis that we're seeing?

Sri Nagavarapu: Yeah. So they did get a chance to see empirical analyses like these along the way. The reliability number was something they were able to see in the proposed rule. And you know, we received public comments, so they are able to see the reliability numbers that you all saw here. For the validity testing they saw measure score distribution, performance by different sorts of complications and so on in a national summary data report as well as in empirical during the problem itself.

(Jack Needleman): And one last question on the 60% reliability consensus standard. This was a development process. So, at what point were there actual votes with the 60% consensus to move forward? You know, was it just the one vote at the end were there lots of votes intermediate? How did those intermediate votes affect the development of the measures themselves?

Sri Nagavarapu: Sure. Yeah, there was a series of intermediate votes at each step. So you could think about that as votes to make specific decisions about the patient cohort. So what trigger codes are involved for incentive? A vote on the service assignment roles, which costs are included in the measure and if there were particular items that came up for discussion and a lot of debate, those specific items would be voted on.

So if there was a particular question about whether a certain type of cost should be included in the measure, then their vote specifically on that, that

was a salient point of contention in the conversation. So there was a series of intermediate votes on each aspect.

(Jack Needleman): Sure.

(Ashley Wilvan): So I think with that, let's go ahead and move into the first measure. And Sri or Nirmal, if you wouldn't mind just giving us a really brief overview of the cataract removal measure in terms of kind of the clinical logic and the approach you used for that specific measure.

Suzanne Pershing: Hi everyone. Good morning. My name is Suzanne Pershing. I'm actually going to be talking about the cataract measure. I'm an ophthalmologist, a cataract surgeon and chief of ophthalmology at the VA Palo Alto healthcare system and residency program director for the Stanford Ophthalmology residency program. And I was a moderator for the cataract episode-based cost measure development through Acumen. So I'll be able to give a bit of an introduction to the measure.

It was developed to identify clinically meaningful services that are related to quality of care, can be reliably influenced by the cataract surgeon and for which the cataract surgeon could fairly be held responsible. And the development process was, as alluded to earlier, led by expert nominated clinicians. In this case, including cataract surgeons who are national leaders and representatives from major societies as well as other ophthalmologists such as retina specialists who are involved in managing cataract surgery complications as well as representation from an anesthesiologist, nurse anesthetists, internal medicine specialists and optometrists as well.

So just to give you all a bit of an overview of the cohort definition or episode window services assigned and risk adjustment, and then also to get into a

couple of high level critiques. First off, the cataract cohort was defined here as routine non-complex cataract surgery, which was set as a CPT code 66984, excluding patients who had a long list of preexisting, mostly ocular co-morbidities, and a rationale for having a limited population was twofold.

First, to ensure a homogenous patient population for a really fair comparison, which is something that subcommittee very much focused on throughout their deliberations. And then also to be consistent with precedent, since the PQRS cataract surgery measures use the same set of preexisting co-morbidity exclusions. And these limitations were something about which the subcommittee of clinician experts really felt very strongly at the outset. So when we began to establish the measure of this was felt to be quite important.

Patients or episodes were then subdivided into subgroups based on unilateral or bilateral. That is, two surgeries performed on one eye and then the other within 30 days. And also by site of service, ASC versus an HOPD. And episodes were sub-grouped by laterality because some services like preoperative exams and testing might be applied to a second surgery performed in close succession and cloaked in the overlapping postoperative period might be difficult to attribute to a specific surgery.

And then we also subgrouped episodes by site of service, HOPD versus ASC, because the attributed clinician may not be able to influence site of service, which has known payment discrepancies. And again, this is something about which the clinician experts felt very strongly. It's an important to note - so we also excluded other populations, I should add. So including Medicare Part C beneficiaries, since data on these beneficiaries was incomplete.

But even with all of these exclusions and subgrouping, the majority of clinicians had over 10 relevant episodes in a given subgroup. Eighty-seven

percent of TINs and 83.7% of PINs and PIs who met a case minimum of ten before the application of all of these exclusions and subgroups still match that case minimum after all exclusions and subgroups were applied, which really helped the subcommittee to feel more comfortable about the fairness and consistency of the measure.

In terms of episode window, the pre- (unintelligible) episode window was selected at 60 days based on empirical data showing that the majority - over 75% of episodes had a preoperative ENM encounter with cataracts as a primary or secondary diagnosis within 60 pre-trigger. And also most preoperative eye testing is usually done within 60 days pre-trigger, just clinically.

And then the post-trigger episode window is collected at 90 days, primarily for concordance with the cataract surgery global period. And also similarly because most episode-associated costs occur within 90 days post-trigger according to the subcommittee.

And then for specific services, the subcommittee was given the ability to choose shorter windows of time. In terms of services assigned, the clinician experts identified services that in the preoperative/postoperative period that could be reliably attributed to the cataract surgeon.

And they actually went through - individually discussed and picked code, which were selected for face validity and included services associated with preoperative testing as well as postoperative complications such as they had dropped cataracts that might require retinal surgery to remove remaining cataract lens fragments, as well as intraocular lens exchange, repositioning or a secondary placement, which indicated a problem with the initial positioning

of the intraocular lens - the primary surgery. And then also things like eye infection and (unintelligible).

The final measure specifications were made publicly available. And each clinician and clinician group alluded to earlier provided a detailed report including the leading costs and clinical themes that may drive into higher low-cost performance. And of course while this is quite a bit of information, it's also completely transparent, which is something we felt very important throughout the process.

Finally, risk adjustment in this case included general variables such as age category, patients being originally disabled, ESRD status, long-term care residence as well as base HCC model, variables and interaction terms. And we also added in episode-specific risk adjustments including a GC modifier to avoid penalizing teaching cases at academic institutions and also for billing new versus established patient ENM code, to avoid penalizing surgeons who have a primarily referral-based practice performing cataract surgery on mostly new patients, since these would have higher costs.

There were a couple of key critiques about the proposed measure that I did want to address directly. And the first was around low levels of variation. I think although the cost of each procedure and consequently the overall IQR for cataracts is small, the variation here is real. And I think this is something that really emphasizes that cataract surgery is about being high volume, even though each procedure individually compared to other things like hospitalization may not rank quite as highly include total cost.

But also the variation is clinically meaningful in a way that can be influenced by clinicians and the difference in cost is substantial and aggregate. Also, important to note that this approach does perform better than alternative

approaches like an all costs measure by A, being directly developed with clinician input to endorse space validity and also by higher reliability, which we measured by a higher proportion of total variance due to between clinician variance versus within clinician variance.

I think the fraction there being closer to 0.95 versus 0.44 in all-cost measure. So it does improve in those two domains. Then the second point is around pass through drugs. And this episode-based cost measure includes a drug called Omidria.

It's an eye injection used at the time of surgery that costs approximately around $450. And it's currently on pass-through status. And this is a lot, of course, relative to the cost of cataract surgery for which I think the professional fees is around $650. And other medications or devices that are sometimes used to address small pupils or pain, for example, are not on past-due status.

So this has been objected to - the inclusion that is - by the imagery of manufacturer and affiliates. But just to point out, the subcommittee did consider this medication and specifically recommended to include it because it's an important area of cost variation to the measure and also relevant to this is the measures focused on parts A and B spending, and in developing this episode, the clinicians were careful to not inappropriately include Part B drugs that may have Part D substitutes.

So for example, intravitreal injections of anti-veg F drugs for postoperative cystoid macular edema, which could also be treated with topical steroids or and eye drops as an alternative.

So I'll go ahead and stop there. We could certainly go into more detail about the measures but if anyone has any other questions. Just to also add, each year with regards to pass through drugs and other developments. Each year we will be doing annual measure maintenance, evaluating new services and technologies and drugs that have become available and making decisions on a case-by-case basis regarding what will be included in the measure, and then these updates being done in consultation with CMS and will include clinician input.

(Ashley Wilvan): Okay. Thanks. Thanks for that. I think what we'll do is just go right into the measure discussion starting with importance. Lena, are you there and are you able to get us…

(Lena Walker): Yes. I am here and I can get you started. I'll start by summarizing a little bit of what was presented by the developer in response to the importance measure report, and then summarize the committee comments from the pre-evaluation survey.

So as the developer just mentioned, this procedure is a high-volume procedure. They've noted that it's the most common surgical procedure in the US including among Medicaid beneficiaries and that Medicare spends more than $3.4 billion annually on the treatment of cataracts. This procedure, cataract extraction with IOL, is the most common procedure.

They provided the mean performance score at both the clinician group practice level and at the practitioner level. And so at the clinician group practice level, the mean score was $3,021 within a quartile range of $238. The spread between the 10th and the 90th percentile was about $584.

The mean performance school at the practitioner level is similar. So the mean there is $3,038 and the inter-quartile range is $232 against similar. Range between the tenth and 90th percentile was $552. The developers' discussion on the opportunity for improvement is - they noted that this procedure has a high rate of surgical success. But there could be complications and where there are complications, it has the potential to result in a much higher episode cost.

The blind complication they mentioned, which I have trouble pronouncing, endosophomitis, has a small incidence - it's an incidence of between 0.05 and 0.2% for Medicare beneficiaries, but it's a - they note that it leads to episode costs that are 83% greater.

So the committee's response to the high-impact, high resource use and opportunity for improvement, there was seven responses to both. And generally all agreed that, it was important to measure. So yes in the high impact and yes in the opportunity for improvement, I would note that as the developer pointed out, I would note that the caveats there were related to the very small inter-quartile range, and so the very small variation.

And in fact in the opportunity for improvement, that was one reviewer who did not find any opportunity, did not think because of the small inter-quartile range that that would be an opportunity for improvement and also noted that there was no data on what's being overused and no information on the distribution of the strata.

And so I think the developer responded to the question about the small variation, but I don't know that the committee has - would want to probe a little more. That concludes what I was planning to say. Oh, well I did want to

add that the committee did - the staff reviewed the information and the staff recommendation was that it was (unintelligible).

(Ashley Wilvan): Thanks, Lena. I wonder if - (Cheryl), do you want to take it from here? I'm happy to help facilitate as well to open it up for the committee for discussion at this point.

(Cheryl): Sure. Thank you and thanks (Lena) for a nice synopsis. So I was the person who called out that the interquartile range difference seemed quite small and I recognized that this procedure affects a lot of people and - but I guess I'm trying to think about this in terms of the potential savings for the Medicare program and would appreciate the measure developer commenting on, you know, whether, you know, an interquartile difference of $173 is actually meaningful given that it's a considerable amount of work for - not only to generate the measure, but for people to have to pay attention to the measure and act on it.

Sri Nagavarapu: Suzanne, you should feel free to jump in. But this is Sri from Acumen. One quick note I'll make. I, you know, I think a consideration, I think our interquartile range is an important one. I think the volume here, the part that that really helps a lot as well as the correlation with sort of areas that Suzanne emphasized could actually be affected. And so on the volume point, if there are around 550,000 reportable episodes, just a savings - the - of $100 on average, on any given episode or $200, translates into - on the order of $50 to $100 million in savings.

And so that’s on the volume. On the other point about the clinical actionability, you know I'll defer to Suzanne on it. I think the correlations that are shown with various service categories like complications are high and

kind of assured us that there's actionability here. But Suzanne if you, if you want to jump in, feel free?

Suzanne Pershing: Sure. Thanks everyone. Sorry, I'm actually en route to the airport now. Yes, and this is something that the subcommittee did discuss too in developing the measure. It was felt that there were actionable things that could be done to improve quality and lower costs. And part of the design of the measure was actually - even though the difference in cost on average are low or having a small IQR, there are real things that we do see, as was mentioned earlier, as substantial difference in cost related to specific complications that we know can be reduced during cataract surgery.

So, despite being low compared to other things that we might look at, I think it is very reasonable to construct a measure around cataract surgery.

(Kristin Martin Anderson): Hi, this is (Kristin). If I could have follow up after you, (Cheryl).

(Cheryl Sandberg): Sure. Go for it.

(Kristin Martin Anderson): How much though of that variation is present once you separate in the strata. Do you see, did you look at the interquartile range by strata?

Sri Nagavarapu: Yeah, this is Sri Nagavarapu. We did. In the testing form, the workbook testing appendix that you all have, that information would be in Table 2B-4-2, score by clinician. What that shows is it has rows for each strata or sub-group. The sub-group is a level of risk adjustment is run, and it shows the interquartile range for each.

The one thing I want to note about those numbers to be careful about in terms of interpretation, is that the numbers that you see there are all scaled and

turned into dollar terms based on the dollars for each subgroup. And that's just the ease interpretation because of something that we've heard from stakeholders about interpreting the interquartile range for each subgroup. But in the actual measure, all the subgroups are - the O over E ratios are averaged together and just multiplied by the national average.

But the inter-quartile range for each subgroup is given on that tab, if people are interested, and sort of mimics the inter-quartile range for the overall measure.

And I will note that in terms of variability we were also thinking - I think it would be - the way that we thought about the opportunity for improvement here is it's something like office visits, just changing a one-off office visit is not where the likely savings is here. The likely savings is in having five or six fewer episodes that have expensive complications and there's very real distinction between providers that are apparent in the measure scores distribution. And I think the reliability number being high helps help make that clear that there's something distinctive about better performance here.

(Kristin Martin Anderson): Thank you. And you, the distribution though seems the prove out something that was (unintelligible) in the upfront measure description, which is that the complication rates are actually quite low. So if you take the volume, you know, the argument that you can take $100 off each, you know, episode and save all this money versus you can take out more money off of a fewer number because of complication, not sure that the volume plus impact components add up, you know, in terms of effort relative to outcome.

Can you comment more on that? I mean, the complication rate - R and D very low, right? Because the big differences between the 1% and the 99%, not really between the, you know, I mean if those complications were more

common, I think you'd see that in a lower - you know, in a higher percentage of the high costs.

Suzanne Pershing Just from a clinical perspective that's true. But we also - I mean one complication is also very, you know, bad and we don't - we want to keep folks from having blindness or serious problems following cataract surgery. So some kind of a quality, an importance to care standpoint, I think it does still make sense. And the complications themselves maybe expensive to manage and treat.

So let's say we had a $1,000 complication. If you can reduce the rate by 10%, would be a substantial improvement, but maybe only lead to a $100 difference, just to throw some numbers out there. But you know, it's the overall cost of each cataract surgery compared to other things we look at in medicine is low.

But I think it's a clinically meaningful and important domain and the, ophthalmology - everything we do is kind of on the lower side of things compared to other specialties that are not involved with inpatient care for patients. And we'd like to have an episode-based cost measure too, instead of being held to Medicare spending per beneficiary, total per capita cost, et cetera. And I think looking at just the real feasible limitations of what we can construct, some of its physician (unintelligible).

If cataract surgery was much more expensive, the proportional difference would probably be greater. But you know, I think that's a subject we wouldn't necessarily want to go in either.

(Kristin Martin Anderson): Yeah. And just a closing comment on that and I'll stop. I agree with you on, you still want to get rid of the complications. I'm just not sure

that a cost measure is the way to do that versus you know, an ambulatory version of a (unintelligible), where you actually measure the compensation directly.

(Cheryl Sandberg): Yeah, this is (Cheryl), I'm going to pile on to (Kristin)'s comment. Yeah, so I really struggled with this measure. I understand, you know, this is a highly prevalent procedure. We're definitely interested in the quality of care delivered around it, but I also struggled with, is a better measure and I feel like I should know whether it's this type of measure exists, if Medicare is really measuring quality or complications rather than cost. So, is there a current measure that deals with the quality of cataract surgery?

Suzanne Pershing: I would just say on the flip side of that too, we do want to have measures that align between cost and quality. So there are measures that are related to, for example, visual acuity outcomes after cataract surgery. And that's highly relevant to quality. We want to have that paired with relevant cost measures. So if we do see things where there's alignment, I'd argue that's a good thing.

(Crosstalk)

Sri Nagavarapu: Real quick, this is Sri Navagarapu. One other number that we thought could be useful for you and in considering this is that, you know, in the interquartile range was showing the provider measure scores. The reason for mentioning something like a $200 interquartile range is that's - showing an average across episodes - it's showing that the costs, even if they're a relatively low probability, less than 20% events or 10% events, that they're large enough that on average they're having large impacts on the average in a way that is indistinguishable among providers.

But you could also look at the episode level and if you look at cost in the first quintile versus episodes in the fifth quintile on average, you see a very large gap in episode costs.

So the first quintile is $2,149. The fifth quintile is $3,728. And so at any given episode level there are things that can dramatically be done to make costs look higher or lower. And what you've seen in the IQR and the measure score distribution is average over the things for providers and their reliability, if I'm making clear, that the very large distinctions in episode costs are not just randomly distributed among providers, but are - there are certain providers that have more expensive episodes occurring systematically.

Suzanne Pershing: Can I ask one more question? Because I know you're stratified by ambulatory surgery centers versus the hospital outpatient department, but it seems to me like that is probably the area where CMS could generate the greatest savings by having a site-neutral payment, potentially over and above reducing the episode costs. Would that be a fair statement?

Nirmal Choradia: So hi, this is Nirmal Choradia. So yes, to an extent that would be a fair statement and I will actually point out across both cataracts and colonoscopy. This was an extremely contentious point with the knowledge - with actually the back and forth of how much effect does an individual provider or a group of clinicians have on being able to practice at an AFC versus an HOPD versus being able to do something in their office.

And you know, it was in fact interesting to hear one of the gastroenterologists who specifically only worked in an HOPD through an academic medical center was actually one of the most vocal people in favor of not subgrouping or even risk-adjusting by site of service. And saying that this would inevitably

force many people to move their practice to an AFC in order to be competitive.

But against that there was also a number of people discussing how they - even across state laws they have very little ability to affect where they practice. And this would also create a wide variation in state by state costs due to just wide differences in the allowability of a freestanding AFCs.

Sri Nagavarapu: And I'll just add to that that, this is Sri from Acumen, this is something that was really extensively discussed in the clinical subcommittees. I know for instance for colonoscopy originally they didn't subgroup by site of service. They had the field testing and the clinical subcommittee members as well as the broader public view the results of the field testing, and the societies in addition to their clinical subcommittee nominees came back and decided that they wanted the subgroup by site of service based on the field testing results that they saw nationally.

So this was definitely something that was closely considered. It is the case that ASC rates are lower in a site-neutral payment. Could save a substantial amount of money. I don't know if CMS is considering something like that on the payment policy side. But on the measure development side it's something that I think would be extremely hard to implement given the perspective of the stakeholder societies that we've seen through the measure development process.

And so we see this as the macro legislation is calling for cost measures. We need to create good cost measures that stakeholders find to be actionable. And so we believe that the cataract measure that the subcommittee chose to start with is one they believe was high impact and a good place to start.

Suzanne Pershing: Okay. So, I'm cognizant of time. I just want to see if there are any other issues that committee members want to flag or ask measure developers before we move on.

(Lisa Last): This is, this is (Lisa Last). I just have a question maybe for NQF staff in terms of criteria to assess this. So we've heard that CMS thinks this is important. We've heard that the ophthalmologists and the clinicians who this would impact think this is important. How do we rate that in terms of our overall assessment? Because it seems like to me, like those are pretty important criteria.

(Ashley Wilvan): Hi, this is (Ashley) from NQF. So we're guided by the NQF criteria. Certainly, those are considerations and I think to the extent that those are a part of your consideration potentially for assessing maybe impact or something, but the criteria that we're looking at today are around high impact or high resource use and opportunity for improvement for this particular criterion.

And so, you know, potentially if, you know, that might be a consideration for whether or not there’s opportunity for improvement or something but if, you know, the criteria don't specifically address support from a particular stakeholder groups as part of the importance consideration. So I'm not sure that there's a great place for that in importance. It will certainly be a consideration as we discuss usability, particularly from, you know, whether it's usable from certain stakeholder groups.

And so we'll discuss, you know, some of the comments we received there and can certainly have a discussion about, you know, the feedback from other specialty societies and so forth in support of the measure.

(Kate): So (Ashley), should we move on to the next step to rate this particular criteria?

(Ashley Wilvan): Sure. If folks are ready to vote unless there's any other concerns to raise or questions for the developer, I think we can call it for a vote and we'll have you guys enter your vote for importance into the Survey Monkey tool that hopefully you have up on your screens.

Woman: (Ashley), just to point out, okay I thought (unintelligible) had his hand raised but he took it out.

(Troy Theisinger): This is (Troy), I'll throw in a quick comment. From the primary care perspective, we have a large number of full-risk contracts on Medicare patients. And by base contract to cost is very important to us, quality is very important to us. So to me in terms of important cost measures are extremely important to us, because we're looking at per member per month on our Medicare beneficiaries, to the tune of tens of thousands of people for our practice.

(Ashley Wilvan): Thank you. And it sounds like we have (Troy) back. Why don't we do this for this first vote? Let's go ahead and use the Survey Monkey and then what we'll do before we vote on scientific acceptability, we'll get a sense of whether or not we have quorum on the call and then we can perhaps skip to the voting tool. Actually, hmm. I'm sorry. I'm thinking out loud because importance is a must-pass criteria it would be helpful to know I think more immediately whether or not it passes so that we can determine whether or not we continue discussion.

So (Troy) you're at your computer and have access to the webinar platform and all the other…

(Troy Theisinger): Yeah, I'm logged in. I'll have access to the rest of the day.

(Ashley Wilvan): Okay. Awesome. So let's do this. Can we do a quick - just a quick roll call to see who's on and if we're on, if we have enough people, I know we went ahead and sent out the Survey Monkey, but we may do a switch back to the poll everywhere pool that we initially discussed in the beginning of the webinar. That way we can either vote on the webinar and decide - we'll know immediately whether or not we move on if something does not pass.

Is that workable for folks or - I don't want to confuse things too much, but it may make our processes a little bit more efficient.

Woman: Yeah. Sounds good.

(Ashley Wilvan): Okay. Let's just do a quick accounting here to see who's on the call and see if we have our quorum. (Navea), do you mind going through the list here real quick? Do you have access to that or I can…

(Navea Kumar): Yes, I can just do a quick roll call. (Cheryl Sandberg).

(Cheryl Sandberg): Here.

(Navea Kumar): (Kristin Martin Anderson)?

(Kristin Martin Anderson): Here.

(Navea Kumar): (Troy Theisinger)?

(Troy Theisinger): Here.

(Navea Kumar): (Rachel Howe)? Okay. (Sonny Gemnani)?

(Sonny Gemnani): Here.

(Navea Kumar): (Jack Needleman).

(Jack Needleman): Here.

(Navea Kumar): (Lisa Last)?

(Lisa Last): Here.

(Navea Kumar): (Janice Orlowski)?

(Janice Orlowski): I'm here.

(Navea Kumar): (John Ratliff)?

(John Ratliff): Here.

(Navea Kumar): (Lena Walker).

(Lena Walker): Yeah.

(Navea Kumar): (Herbert Wong).

(Herbert Wong): Here.

(Navea Kumar): And (Trini Basudara), are you able to join us? And (Ashley), it does look like we have quorum so we are able to vote.

(Ashley Wilvan): Okay.

(Brent): (Brent) (unintelligible) joined, just...

(Ashley Wilvan): Oh, hi. Hi, (Brent). Welcome. What a pleasant surprise. Welcome. So I think we are good to vote. So let's go ahead and shift to the poll everywhere school. (Navea), if you want to kind of orient everyone again to what we'll be doing and we'll go ahead and do the first vote on importance when we get - when we all get there.

(Navea Kumar): Sure. So we do a voting, a test flight just to make sure everyone is able to vote?

(Jack Needleman): This is (Jack), I don't have access to the poll everywhere tool.

(Navea Kumar): (Jack), I can send you another email.

(Jack Needleman): Terrific, okay.

(Ashley Wilvan): I promise it will be efficient when we get there.

(Navea Kumar): Sorry, give me one moment.

Woman: Okay, so all we should see right now is this blue screen, right?

(Navea Kumar): Yes, that is correct.

Woman: Okay, thanks.

(Navea Kumar): Okay, (Troy), I just sent you the instructions and I will activate the poll for test. One moment.

(Troy Theisinger): This is where I've got the blue screen up, so I think I'm at the right place.

(Navea Kumar): Perfect. So you should all…

(Ashley Wilvan): I think (Jack) needs the link. I just wanted to make sure we sent him one as well.

Man: I need the link as well, (Brent) (unintelligible). I'm sorry, I'm on the webinar, but I'm not seeing it.

(Navea Kumar): Sure, I can resend it to you. And for those who are able to have the link and access the voting poll, the test has been activated.

Woman: Great, yep, I see it.

(Navea Kumar): Okay. (Jack) and (Brent), I just sent you the voting instructions once again. Please let me know if you're unable to get to the link.

(Brent): For clarity, this is not the Survey Monkey tool. This is, because I clicked on the link and I have the question in front of me.

(Navea Kumar): Yes. This is the poll everywhere. It's a different platform.

(Brent): Okay. I'm not on that platform so I'll wait for your email.

(Cheryl Sandberg): (Brent), it was in the email sent to the committee, yesterday I believe.

(Brent): Thanks, (Cheryl).

(Cheryl Sandberg): If you can't find that, let me know.

(Brent): I got it. I got it now.

(Cheryl Sandberg): Okay, cool. Okay.

(Navea Kumar): Okay, great. It looks like everyone is able to…

Man: I've got access to the vote everywhere from the link that was sent.

(Navea Kumar): Perfect. And with that, (Ashley), shall we begin voting for importance?

(Ashley Wilvan): Yup. Let's go ahead.

(Navea Kumar): So the voting is now open for importance to measure and report. Your options are A, high; B, moderate; C, low; and D, insufficient.

Looks like everyone voted. Voting is now closed for importance for measure 3509. The results are two votes for high, seven votes for moderate, one vote for low, and one vote for insufficient. The measure does pass with nine votes.

(Ashley Wilvan): Thanks, (Navea). So we will keep going. We'll move into the scientific acceptability criteria. And I know that again, (Triavas) was unable to join us today due to a last-minute conflict. But if (Herb) or (Lisa) have any thoughts that they could share with us to get us started, otherwise we could do another - we could do a brief intro from the facilitators.

(Lisa Last): This is (Lisa), I can talk about reliability.

(Herbert Wong): (Lisa), we can - I can tag team with you, so you can get us started and if I have anything to add, I will contribute as well.

(Lisa Last): Okay, all right, sounds good. So (unintelligible) actually talked about a lot of this and sort of the prep for the measure. And as a reminder, all three measures used a similar methodology for liability and reliability testing. So, this is a Medicare measure and the testing was done on a Medicare beneficiary, TIN-level testing. They used a test/retest method with correlation and signal to noise.

And using these methodologies - the test/retest, the coefficient - the correlation coefficient was quite high between clinician groups and clinicians. .93 and .92 respectively. And then they looked at the test/retest, looking at a variety of clinicians and the groups and the lowest quintile spending in the first sample, 70% of the groups and 67% of clinicians in the lowest quintile - lowest spending quintile in the first sample were also in the lowest quintile in the second sample.

So there was a high test/retest correlation in that - in 92% of groups and 89% of clinicians were in one of the two lowest. So using these methodologies, there was a fair degree - a fairly high degree of consistency and it looks like a fairly high degree of reliability. I don't know what this means that they used to Adam's method for signal-to-noise analysis. So somebody who's more familiar with that could speak on that.

But they found in that method the reliability score was also high, .95 for groups and .94 for clinicians. And they use TINs and PIs to distinguish between those two.

Looking at the staff rating, the staff rated this as moderate. And looking at the comments from the people who commented ahead of time, most of the comments were good. There was one concern about the reliability score of .4 for the TINS, which is considered a low reliability. And then there was another concern about a difference in reliability between a 10 case, 20 case and 30 case minimum, which is greatest for the tenth percentile and desire to hear more about that.

So based on my review again, not being a methodologist here, but it looks like assuming this was a valid testing methodology, which it would seem to be, I would say that the reliability seems solid. Anything you want to add?

(Herbert Wong): The only thing I would add to that, (Lisa) that was a terrific summary. So thank you very much. The only thing I would add was a couple of things that I noticed in terms of the commentaries on reliability. Oftentimes we do drift from looking at reliability and validity in terms of some of the measures, but in terms of the specification, there was a comment about whether or not the opinions of physician experts in terms of exclusions for comorbidities - that may have been addressed during the initial characterization by Acumen about how they have established that. So, that might be cleared up.

The second point is that there was - a reviewer did notice that there were differences between the different cases that were being evaluated. So a case that had - an observation that had 10 cases versus 20 cases, 30 cases, et cetera, had some variability in it and whether or not that would impact the measure's reliability and some commentary from Acumen might be helpful there.

(Cheryl Sandberg): And this is (Cheryl), just to add on to (Lisa)'s comment about the Adams method. Essentially what that assessment is looking at is the ability of the measure to discriminate performance between providers. So, generally the community thinks a measure that's .7 or higher on a scale of zero to one have acceptable reliability. And reliability is up in the .95 range signal, very high ability to differentiate performance between providers.

(Lisa Last): Thank you.

(Jack Needleman): Yeah, this is (Jack). If I can just follow up on (Cheryl)'s comment. The scientific methods committee spent a lot of time thinking about the standards for reliability that they would feel comfortable with. And as (Cheryl) noted, somewhere between .7 and .8. No clear - consensus did not emerge as to which end of that. The statement that CMS use .4 is a moderate threshold. To be blunt, the .4 threshold did not feel comfortable to the scientific methods committee based on the vote.

But of the three methods measures we're considering today, the reliability test, whether it signal-to-noise or test/retest or the Pearson correlation on a test/retest, this measure had the highest rung liability of the three measures we're looking at today. And I don't think based upon the testing the committee had a concern about that.

(Cheryl Sandberg): Thanks (Jack) for clarifying. And I would underscore that reliabilities below .7 are generally considered to be conveying a lot of noise and not much signal. But this measure does not appear to be suffering from it, nor do the other two measures that we're considering today.

(Ashley Wilvan): So this is (Ashley) from NQF. I just wanted to point out a couple procedural things with the voting. So this was one of the criteria that was evaluated by the methods panel. They did provide a vote. The committee, I believe the vote, the next vote that we will have for you guys is to vote on whether or not you'd like to accept the methods panel vote on reliability, or to be able to vote and apply your own vote to that - to this criteria.

So if - you know, certainly we can have more discussion about this, but if folks feel pretty comfortable with the reliability here and the methods panel's assessment, we can call up the voting tool and have you guys vote on whether or not to accept the methods panel vote. And I will just point out, (Yak) as a member of the methods panel, you do not need to vote again on reliability or validity.

(Cheryl Sandberg): (Ashley), do you want to just quickly summarize what the methods committee recommended?

(Ashley Wilvan): Sure. So, their vote for the reliability testing and specifications was a moderate. And so generally they, just like (Jack)'s summary, basically the committee felt very comfortable - I'm sorry, the methods panel felt very comfortable with the correlation coefficients, the stability across the quintiles as well as the signal-to-noise ratios and the Adams R score as well. So, that is reflected in their moderate grading of reliability.

(Cheryl Sandberg): So I'm comfortable with accepting the scientific review committee's recommendation here. Are other members of the committee?

Woman: I am.

Man: Yes.

(Ashley Wilvan): (Navea), if you want to pull that up and we'll just have them vote on that and then we'll move forward to validity.

(Navea Kumar): Sure, (Ashley). Okay, voting is now open for measure 3509, takes up the scientific methods panel rating for reliability. Voting is now closed. All members vote to uphold the SMP's rating for the reliability.

(Ashley Wilvan): Thank you. And so just a brief time check. We are scheduled to adjourn for this morning's session in about ten minutes. We do know that the first measure does usually take the longest. I wanted to just open it up to the group here to see whether or not you'd like to use these last ten minutes - we dive into validity and pick up where we leave off at three. And, or if there's other proposals for how we use the last ten minutes of the call.

Sri Nagavarapu: And this is Sri Nagavarapu. Sorry, just as a quick note, our ophthalmologist Susan will have to leave after the - these 10 minutes because of another commitment. And so, if there are clinical questions, that would be good to get her thoughts on, that'd be great to use this time for that.

(Jack Needleman): So, this is (Jack). One of the issues that the scientific methods committee kicked down the road on validity was the appropriateness of the exclusions, which given that 50% of the cases, 40% of the cases get excluded with the exclusions. I'd like to have some discussion of validity before we lost the ophthalmologist.

(Ashley Wilvan): I think that sounds like a fair approach.

Suzanne Pershing: Well, if you'd like me to just chime in a, this is Susan again. So a couple of points about the exclusions. There are a lot of them and it was important to the

subcommittee to make sure that we didn't lose a substantial number of clinicians, TINs or TIN NPIs who would have otherwise been able to meet the - meet episode minimums.

So we did find that over, I think 87% did still maintain episode minimums after both exclusions and subgrouping because it'd be looking at the exclusions across each of these groups. So site of service with unilateral or bilateral cases. The reason for choosing such a large group of exclusions was to make sure that we had a homogenous episode group as much in this patient population and to make sure we were consistent with the PQRS measures.

I will say just from a practical standpoint, it was important to have a measure that would be clinically acceptable to - especially for a first wave - to the population of providers that would be affected by it. And almost the first comment from the clinical subcommittee was, we need to make sure we're excluding all of these conditions because we're already doing this for other measures.

We want to be consistent and there are things around types of - more complicated cataracts for example, certain types of glaucoma like pseudoexfoliation will make surgery more challenging, more likely to have complications, and would be something that they felt they couldn't adequately risk-adjust for.

Some of these patients might be more likely to have a different CPT code for cataract surgery at 66982 instead of 66984 that we weren't capturing. So for a first wave, especially since we were still able to capture the vast majority of clinicians and not lose them by episode minimums, they wanted to take this conservative approach. Hopefully that gives a little bit of background.

(Jack Needleman): Great. I'm looking at the documentation while you're talking, trying to multitask here. And, you note here that the use of the ocular condition exclusion are consistent with two NQF-endorsed quality measures on outcomes. So, does that mean that we're basically comparing the cost of the - when we do that standard grid of cost versus quality that CMS does on these measures, does that mean we'll be comparing basically the same cases? The cost measure and the quality measures of both basically on the same sets of cases.

Suzanne Pershing: You have it exactly right. The idea to have consistency across the population. Again, aligning cost to quality. Did you guys hear that?

(Jack Needleman): Thank you. That resolves all my concerns about validity.

Suzanne Pershing: Thank you.

(Ashley Wilvan): So, we did have a lead discussant assigned. (Sonny), are you there?

(Sonny Gemnani): Yes, I am.

(Ashley Wilvan): I wonder if you could just give us a real quick summary of some of the things you identified in your review to key up the committee in terms of some other issues that might be - that should be discussed.

(Sonny Gemnani): Sure. Firstly, the methods panel, as you know, did not reach a consensus. There were three people who had moderate and two who had low, graded low. In terms of key points aligned, I think the main problem that I saw was in terms of exclusions, which we just talked about, several (unintelligible) members and SMP members had some issues about exclusions, which I thought were valid and we just addressed a few moments.

In terms of the empirical testing approach, I did notice that the methods were pretty much similar across all three measures in terms of dividing them into clinical themes and looking at their correlation. And the numbers varied from zero point - at least for cataracts, varied from .18 for cataract surgery related office visits to .51, which are complications.

The - our score were if I remember was pretty robust. I think .75 if I remember off the top of my head, for that. For other things which I think were pertinent, I think pass-through drugs, which I think we had one letter from the American Society of Cataract and Refractive Surgery that we - I think this was discussed at the start of our conference call. And unless some of the other members have any issues to talk about, with that I think I'm pretty satisfied with what I heard from the developer with that.

In terms of other issues, I think that a couple of members from our committee had some issues about social risk factors and ocular co-morbidities and complications and looking at the robustness of the correlation with the clinical themes, ranging from as I said from .17 to .5, and that not being strong or at the best being moderate to low.

That is something that we can discuss now or later. And the final issue which I had concerns and which was more appropriate for section one was how much a difference is - substantial difference in terms of IQRs. But, this has already been covered in one. That's it.

(Ashley Wilvan): Thanks (Sonny). I did want to have an opportunity for a brief public comment period before we adjourn this morning. I know some folks joined this morning with the hopes that we would have gotten maybe a little bit further, but that's okay. We're making good progress. So I'd like to give an opportunity to those

on the phone who wanted to comment on this measure before we adjourned. We re-adjourn for three o'clock and continue with validity for this measure. Just of note for those commenters.

So (Navea) at this time, do you want to, open it up for comment?

(Navea Kumar): Sure. Thanks (Ashley). So at this time, we'll open the call up for any public commenting. If you do not want to make an up public comment, please place yourself back on mute.

(Ashley Wilvan): So we do have a hand raised from Allison. If you have a comment, you can go ahead and speak.

Allison Madson: Okay. Thank you. Can you hear me?

(Ashley Wilvan): Yes.

Allison Madson: Okay. Hi, this is Allison Madson with the American Society of Cataract and Refractive Surgery. I did just want to address that issue related to pass-through status and note that, you know, our concerns related to it aren't specific about the, you know, the clinical relevance of the one particular drug that's (unintelligible) included in this measure at the moment. It's more, you know, philosophically with the idea of including any pass-through drugs, not only in this measure but in any episode-based measure.

For those who are not aware, CMS has created this status for new drugs that come onto the market that are administered during surgery. They are not bundled into the facility fee. They have about a three-year pass-through process, and during that time position, you know, can use them. They're not, bundled into the facility. CMS pays for those separately under Part B.

During that time, CMS looks at the usage of that drug and then following the pass-through period, readjusts the facility fee to reflect the usage of that drug over the first few years on the market, and that is now bundled into the facility fee, you know, in subsequent years.

The idea of the pass-through period really is to sort of have that on uninterfered time to have doctors get used to using the drug. So, we're just concerned that these measures would have an unintended consequence of influencing the usage of those drugs not based on a clinical need, but more specifically just based on performance on the measure. And so we're just asking that they be excluded, and not recommending that NQF endorse the measure until any pass-through drugs are removed. Thank you.

(Ashley Wilvan): Thank you Alison. Aare there any other public comments? We're right at the top of the hour, so we'll try to stick to that time as close as possible, but if there are any other comments and you can be brief, we'll hear that now.

Man: And just in case they're not, Suzanne is still on the line and could respond quickly to the public comment if that'd be helpful.

(Ashley Wilvan): Sure.

Suzanne Pershing: Great. Hey, everyone, hopefully you can still hear me, I'm on board a plane now. So, thank you Allison. I think - we of course had discussed imagery specifically during the measure development process, at the time being the only pass-through drug that was available related to cataract surgery. And for the reasons I've mentioned earlier, the subcommittee collectively had felt that it was appropriate if we took into account clinician input, ASCARS was involved in that as well as PAs or other organizations.

But you're absolutely right as far as going forward with regards to like pass-through drugs and other new developments, not just drugs but also new devices, new techniques, and a new procedure code that may become available. It is important - and each year we will be doing annual measure maintenance.

So continuing to evaluate these new services and making decisions on a case-by-case basis about what to include in the measure. And importantly, I think the subcommittee too wants to take into account, since this is focused on parts A and B and making sure we're not unfairly incentivizing or disincentivizing based on Part D medications. So taking all of those factors into account, but I would encourage the NQF not to not endorse the measure just because it will need annual measure maintenance and updates, which is something I think we all recognize as being important.

(Ashley Wilvan): Thank you. Thank you both for that. So we are, we did pass a little bit on - I want to be respectful of folks' time so that, you have enough time to attend to your other priorities before we reconvene at 3:00. And with that, I think we'll go ahead and adjourn. (Cheryl), do you have any parting comments before we adjourn until 3:00?

(Cheryl Sandberg): No, I do think the conversation has been very helpful and, you know, obviously we're going to work through the other two. We may need to pick up our pace a bit. So, anyway, I think that's it.

(Ashley Wilvan): Okay. Thanks everyone. We'll reconvene at three o'clock.

(Cheryl Sandberg): Thanks, (Ashley).

(Ashley Wilvan): Thanks, bye.

END