NATIONAL QUALITY FORUM

Moderator: Benita Kornegay Henry
June 27, 2019
3:00 pm ET

(Ashley): Okay. So I'm going to go ahead and get started. We only have about two - we have about two hours to hopefully get through the voting for the measures and hopefully we'll get - make sure that we have quorum for this call before we get started. So we're going to start out with a brief roll call to see who's on and we will figure out a plan for voting based on who we have on the phone.

So with that, I'm going to hand it over to (Nadia) and (Kaye) to do a quick roll call.

(Nadia): Thanks, (Ashley). So (Brent Asther)? (Sheryl Jenner)? (Sheryl), you're on mute. Okay, we might have lost her. (Christine Martin-Anderson)?

(Christine Martin-Anderson): I'm here. Thanks.

(Nadia): Thank you. (Troy Pfizenger)? (Nancy Good)? (Andrea Galvern)? (Leighton Hao)? (Sonny Vilnany)?

(Sonny Vilnany): Present.

(Nadia): Thank you. (Lisa Lock)?

(Lisa Lock): I'm here.

(Nadia): Thank you. (Jason Lock)? (Jack Needleman)?

(Jack Needleman): I'm here.

(Nadia): Thank you. (Janice Orlowski)?

(Janice Orlowski): Present.

(Nadia): Okay. (John Gatnes)?

(John Gatnes): I'm here.

(Nadia): Thank you. (Ginnie Lastridar)? (Lina Walker)?

(Lina Walker): I'm here.

(Nadia): Thank you. (Joe Weinjob)? (Herb Wong)?

(Herb Wong): Here.

(Nadia): Did I miss anyone? Or has anyone dialed in while I was calling roll?

(Sheryl Jenner): (Sheryl). I'm still on the line.

(Nadia): (Sheryl). Thank you, (Sheryl).

Do we have - no I think. I think that makes us not in number. We're not going to reach quorum. Would you like me to send out the (unintelligible), (Ashley)?

(Ashley): Okay, thanks. Hopefully we'll get a couple more folks that will join us. But for now, again, if you have access to the email that (Nadia) sent out as we were beginning the morning webinar with the SurveyMonkey link, if you could up that again, we're going to plan to vote via SurveyMonkey for now.

If you need us to resend it, we can do that too and maybe we'll just do that for everyone just in case. (Nadia), so you want to resend that.

(Nadia): Sure.

(Ashley): So we left off talking about validity for Measure 3509. Before we left, (Sonny) gave us a brief overview of some of the issues that he'd identified as we (discussed it). We had a comment from (Palison), from the American, as I have her - from the American Society of Cataract and Refractive Surgery. And also the Acumen Team responded to that based on concerns with the inclusion of pass-through drugs and the counter cost for this measure.

So I did just want to check in before we dive. Again, Acumen Team, are you there?

Man: We're here, (Ashley).

(Ashley): Okay, great. Thanks.

So why don't we do this just to make sure we start off all on the same page. (Sonny), do you mind just giving us a really quick overview again of validity so that we can kind of get our mindset back to where we left off?

(Sonny Vilnany): Sure. So thank you for coming back.

So just as a brief overview. Firstly, we're looking at seven sub-criteria within the topic of validity. Just to refresh everyone's to mind, the methods panel did not reach consensus and it is up to us to come up with consensus on the validity rating.

The (N2S) stat also did not give us any guidance on that based on the S&P's rating. Looking at this seven sub-criteria within validity, measure intent was I thought people have pretty much agreed on everything.

The other issue was about validity testing. As you know that the measure did not meet (N2S) phase validity testing and therefore had to resort to empirical testing which was done by correlation analysis which have correlation of 1.04 compared to 0.95 with complication. This was followed by correlation analysis for clinical themes which is consistent for all the three measures that we are looking at.

The clinical themes for IOL which is the measure that we're currently looking at, the five clinical themes, our (CAD) regulated office business complications, office-based diagnostic testing, office-based procedures and other ancillary care. And the correlations vary from 0.17, the lowest, which is office-based diagnostic testing, to a max of 0.51 which is complications.

As (Jack) had previously referred, previously the S&P have some concerns about the validity testing and access to look more into it. The (arts ware) was,

for the testing, was 0.75 and I think looking at the predictive ratios of the model, it turned out to be one for adjusted and non-adjusted. So those were the numbers that we got.

In terms of inclusions and exclusions, I think we had a lot of comments from different stakeholders based on appropriate inclusions and exclusions. We briefly touched on that in our last call. So that was about that.

And then the final thing was about meaningful differences. And I think we touched mainly on that basically on how much is the meaningful differences based on percentiles or (IQRs) or (debt) files and (it's quite) tier one. So that was reached.

In terms of some of the comments made by our co-members in this committee, I think the major things that people were concerned about is number one, in the testing model looking at correlations to complications where some people did not feel that the correlation was strong enough and at best moderate. Some people had questions about inclusion and of social factors and (SCS) factors and the robustness of the model and the limitation of the current method in addressing that.

And the other concern was pretty much within the IT war range which I think we covered. So that's pretty much a bright, pre-yield overview of that - everything that I have for validity.

(Ashley): Thank you, (Sonny), for that.

I wonder did the other folks who had validity wanted to share some comments. (Troy), I'm not sure if he's on the phone. (Jack), have you had any comments to add in particular? I think it might be helpful to share with

the steering committee the nature of the discussion or the struggle that the method panel had and coming to consensus on validity around this methodological issue.

(Jack Needleman): This is (Jack). I think a lot of the concern or discussion in the steering committee around validity had to do with, excuse me, the appropriateness of the correlating cause with components of the cost basically with the complications or with patients who had gotten additional services. This cuts across the three measures we're looking at today.

I am less concerned about that than some of my colleagues were on the committee. I think those correlational approaches with cost components and the things that you expect to drive up cost where they might be signals of poor quality seems to me the perfectly reasonable way to approach the analysis for a new measure.

(Troy Pfizenger): And this is (Troy). The point I was trying to make on the correlation is it's in moderate, not strong. I do agree empirically complications make senses on the - that cost of this (unintelligible). If makes sense if that cost, whether stronger, is the diagnostic testing and cataract surgery laid off because it had poor correlation so in effect cost much. (They did see) the slightly orchestrate the case a little bit but I don't think it's a major issue.

(Jack Needleman): Yes, I agree. I think this case was slightly overstated.

I wonder if any other committee members have any issues about the testing approach for validity.

(Sheryl Jenner): This is (Sheryl). I have a question that it (unintelligible) all three measures. I thought maybe now is the time to have it although I think kind of where most important is on the measure.

So it would be helpful if the measure developer could clarify (SCS) work that you did to explore the role of special risk factors. What I didn't see is the write-up was at a conceptual model. And it seemed as though you're getting significant results although the effects on this measure particular with or without the risk adjustors didn't seem to have mattered much.

But I was kind of thinking ahead to the knee replacement measure, trying to understand the thinking behind the (SCS) adjustment and why not include those variables particularly for that other measure and whether you have a conceptual model.

(Alex Sandio): (Sheryl), thanks for asking that. This is (Alex Sandio). I'm a cardiologist here at Acumen. And myself and (Trey) will answer that question together.

We have with - if you see measures along with the other five, all of these measures looked at social risk factors and looked at how social risk factors affected measure performance, both in terms of the risk adjustment model, the measures for them and maybe most importantly, physician performance and do that in a systematic fashion.

So what we did is we added the following variables to our risk adjustment model. Raised from the (medical enrollment file) text, (dual status) which has been shown to be a very strong surrogate versus (unintelligible) (status), as everybody knows. And then additional neighborhood characteristics based on a patient thought group, the smallest geographical unit.

So if the (unintelligible) publishes a sample data somewhere between a few hundreds to a couple thousand patients, the blank group characteristics that we added were income level, education level and unemployment level.

So what we did was we took a base model that we started this for each of our cost measures and then we added in each of those variables, first kind of (T slide), adding a couple at a time and then eventually adding all of them in a model.

And then what we found was first in terms of the risk adjustment model, adding all of the variables led to very little change in terms of the (arts ware) of the model. They're kind of (total discrimination). Then kind of diving a little bit deeper into this (unintelligible) model, we looked at the coefficient directions and magnitude. They vary pretty substantially, both between different measures as we might expect but also the changes in sub-groups within the same measure. So fairly inconsistent.

But I think most importantly when we think about this is how does adding social risk factors affect clinician performance. So we took the clinician's percentile rank with the base model that we use in each of these measures and then we compared that to a clinician's rank if the risk adjustment model includes his entire set of social risk factors.

And what we found was it led to very small change in clinician percentile ranking. (Mix) as a program judges the cost measures based on docile benchmarks and so thinking about each of these three measures that you brought up and I'll (focus on the) knee arthroplasty is the one that you face kind of the most central concern on. We looked at how many clinicians actually have a change in their percentile rank by more than 10%. And knee

arthroplasty it was 2.6% of clinicians that had a greater than 10% absolute change either up or down in their (Mix) percentile.

For cataracts it was about 0.15% and for colonoscopy it was less than 0.1% of the total clinicians who were greater than 10 percentile point change in their absolute (Mix) percentile. So when we looked at adding in those social risk factors, we found that it had a very small effect on clinician performance on the measures. And we're happy to kind of answer any additional questions about that.

(Sheryl Jenner): So I think you're right for the first two measures, we're going to consider but for the knee measure, I guess in particular, did you examine the - within provider differences while controlling for - or allowing to show between provider differences on this measure?

Because I'm guessing that probably the risk adjustment matters for about, you know, small fraction of providers who have a disproportionate share of the higher risk patient. And I don't think it's uncommon for different risk variables to move in different directions. So that, you know, result didn't surprise me.

But I think I was sort of struggling a bit to try to understand what the effects are on providers who have a very large concentration say at (duals).

(Trey): Yes. This is something we were concerned about as long - this is (Trey) from Acumen. And so the way we thought about that question is, you know, we saw the positive coefficient for (duals) in one particular sub-group, the knee arthroplasty measure. And so our concern was, you know, I think the same as your concern that if you have a certain provider that have a large share of (duals), then there could be a systematic effect on their provider.

So what we did was translate the provider scores into the percentiles of the (system) as they would go into (Mix). And then look to see whether there were even in the tail of the distribution, how many providers were actually affected because the providers that would be affected would be the ones who had larger concentration to (duals).

Unfortunately, we see a very strong correlation of the measures. I think around 0.989. Between the measure with and without socioeconomic factors. And we also see that the fraction of providers with the move of 10 percentile pointed very small.

And so, yes, we had that same concern. And I should also note that Dr. (AJ Yates) who is on our clinical sub-committee and chairs the clinical sub-committee for the knee arthroplasty measure is also on the line.

Dr. (Yates), if you want to jump in, feel free.

((Crosstalk))

(Ashley): Hi. Before you do that…

(AJ Yates): Yes. Hello. This is…

(Ashley): Hi. This is (Ashley). Hold on. I just want to pause. We're talking about the eye measure now. And I know (Sheryl) had a question about the knee measure. But we could just hold that comment until we get to the knee measure. I think that will be really helpful making sure we can get through all three measures for this call.

(AJ Yates): I'm a practicing clinician, and I'm in a middle of about four different things and I've already spent an hour and a half earlier today.

Just let me tell you something very succinct and very quickly. The reason - I live in a world that's taking care of a high distribution of patients that are at socioeconomic risk. I'm a Neurocity Hospital surgeon.

The reason this is not as bad as you would think in terms of being a risk factor is because Acumen has done a good job in terms of the cost allocation or the assignation of costs. And is the assignation of costs that protects, for instance, myself, somebody who is coming back to the ER and sent to the office, a lot of the cost assignation ends up going to the quick people that are ordering things in the emergency room or somebody that orders something for a follow-up visit that wasn't necessary.

They may not seem impure to anybody out there but it's very different than say the assignation with complications and other measures for hip and knee which are absolutely going to be mine. So these things that I cause like a hip infection or a dislocation or something to that effect, that absolutely gets assigned to me but the stuff that goes on in the margin because of the socioeconomic urban topology that I live within gets muffled by the fact that the cost assignation has been carefully looked at.

So I wanted to throw that out there because it's applicable to the other measures but - and I'm just afraid I might not be on the phone later so I want to make sure I get that in. I'm sorry to interrupt.

(Ashley): Thanks for that.

So if we could just redirect a bit, I just want to make sure we can get through this measure. Working talking about the validity for the ophthalmology measure and particularly their concerns about risk adjustments with that measure. We'll try to stick there and then we'll move more systematically through each measure going forward.

So do other committee members have concerns about the validity of the measure that they'd like to raise?

(Sonny Vilnany): This is (Sonny). There is one reviewer from the S&P who had a consistent comment across all three measures that we're reviewing. And the comment pretty much goes this way, states that this model includes a number of correlated variables. The impact of multi-co-linearity is not addressed in the summary and could be creating anomalies in the model.

I was wondering, (Jack), how was this addressed and Acumen, what was - what were your thoughts on this S&P's reviews and comments?

(Jack Needleman): Okay. I don't think it was addressed and was not fully discussed in the S&P and I think when we have our next meeting we'll have to come back to it. In the core risk adjustment model, I don't care about the co-linearity. I don't care if there's some error in the individual estimation of individual coefficients because there's not a real effort to throw things out.

On the (SCS) stuff where the comments made about inconsistent coefficients of collinear variables, I do think it's a bigger issue and that's why I think that sort of thing that Acumen did which is to look at the relative movement of providers when the variables are in or out addresses the consequences of having these variables in the model.

And that to me was more informative and I think for me at least, the lack of movement across the - with or without the (SCS) variables is compelling in terms of not including them. I think there's another thing that goes on with the (SCS) variables. I think there is a lot of concern and legitimate concern about the way these payment systems treat high minority and disadvantaged institutions and individual providers that are treating large numbers of minority or other disadvantaged populations.

But I think one of the reasons why we don't see the effect in the risk adjustment patient level is even the disproportionate serving institutions serve on minority of those populations. So you see them spread out and the risk adjustment effects spread out across a large number of providers even - and therefore gets diluted.

I do think at some point NQF then have to come back to thinking about the sources and implications of lower - what often are lower quality rankings for high disadvantaged serving institutions. But I don't think what we're seeing in this data is that the risk adjustment model at the individual patients level is not the way we're going to solve that problem or deal with that problem.

(Sheryl Jenner): (Jack), this is (Sheryl). You know, I agree with you. I think it was reassuring to see that, you know, with or without the variables the scores didn't change much, you know, particularly for the first two measures. So thanks for underscoring that point.

(Jack Needleman): Thank you.

(Lina Walker): This is (Lina). (Jack), thank you very much for that explanation. That really helped clear some of the questions I had in my mind. And I did think it was

very compelling that the scores didn't move much with and without their (SCS) variable. So thank you.

Man: All right, I just want to bring up to the committee to see if they have any more questions on validity before we - so that we have no questions in our minds if we start voting on validity.

((Crosstalk))

Woman: Sure. I just have one question. I know, you know, before the break there were - we have a question from the commenter regarding the exclusion or I'm sorry, the inclusion of pass-through drugs and the developer commented on that. That is also something that was discussed in the (top) discussion particularly related to clinical exclusions.

So I just wanted to see if the committee had any thoughts on that because I know that this is an issue that has been pervasive in this - and discussion of this measure within the ophthalmology circles.

I'm not sure if the - is it clear what the issue is or is the committee - is there no thoughts on it? I guess I'm just trying to get a sense of what the stand is so that there could be some clarity for the commenter as well as to how that's being considered as part of validity.

(Troy Pfizenger): This is (Troy). Can you restate the question to make sure I have it clearly on my mind?

Woman: Right. So if you recall in the morning session, there was a commenter that raised an issue regarding the inclusion of cost for pass-through drugs which

are basically a status that CMS assigns a certain drug that are newer to the market to see how - what their use is and so forth.

And so there are some concerns, particularly with the commenter, that (unintelligible) in one of the call from the American Society of Cataract and Refractive Surgery that (deep cuts) are included and she was expressing concern that perhaps these codes or that cost should not be included in the measure.

There's also a letter from her starting on Page 28 of the packet or from her organization on Page 28 of the packet that discusses that issue but it relates not only to reusability but also a bit to validity as well.

(Troy Pfizenger): So this is (Troy).

((Crosstalk))

(Troy Pfizenger): My opinion is that per cost should be included. It's being used. It's being billed. In (unintelligible) contracts it's going to have to be paid for.

(Sonny Vilnany): This is (Sonny). I think the points made by the Society of (unintelligible), there's one thing that I noticed is that there were 10 ophthalmologists on the panel, if I remember, from Acumen Clinical Committee on the list and we browsed on the subject initially during our first glance through how - if the society even endorsed it.

When you talk - when you did - this is the developer, when you talk with the physicians and you feel something (unintelligible) within the group, did you get a sense this was a big, huge issue from talking to all the ophthalmologist

or the magnitude of this problem was that the (team) should be like coming from a niche group of people?

(Trey): This is (Trey) from Acumen. This was something that was not a big issue during the measure development process. So I think (Suzanne) walked through the deliberation to the clinical sub-committee on this. This was a cost for the ones - the pass-through drugs that a major (unintelligible) that's steering out which is used by a very small question (unintelligible) on the order of 6%. The 6% of that (unintelligible).

This is an issue that came up in the clinical sub-committee and people decided to include that cost in the measure. And the one thing I will note is that going forward, as (Suzanne) mentioned in response to the public comment, we absolutely think that each time a new drug comes through and has pass-through status temporarily, it should be considered on a case by case basis whether it makes sense to include it or not.

And we recognize that not all of the new drugs that come on the market should necessarily be included and there are important clinical considerations there. But for the drug that was included, that's something the clinical sub-committee thought about and we have heard from the industry about this. But it's not (lifestyle) concern that we're seeing.

Man: Thank you.

Woman: So it's a decision criteria that you're going to use going forward as whether there's an alternative option that can be used instead of the pass-through drug. I guess I'm trying to figure out how that decision will be made in the future.

(Trey): Yes. The what - the way we're thinking about this is there's a couple of considerations. One consideration is the extent to which there are Part D drugs that are substitutes for that drug.

So to give you an example, currently Part D drugs are not included in the measure because there's costs standardized and Part D costs can be all over the place. And our (staff) was very clear about Part D costs and if their standardized.

And so what we want to do is if there's a new drug that comes in, once they're on pass-through status, if that is a substitute for Part D drugs, we would lean towards not including that in the measure because the idea being that if you include that in the measure then you're sort of bias measure against beneficiaries who don't have Part D enrollment.

And so that's the case where we would want to be extremely careful about including those drug costs in the measure. There are also other cases where we could imagine cases where there's a pass-through drug that affects certain sports of outcome that are hard to capture that CMS may not want to incentivize. And so there'd be deliberation on a case by case basis for those sorts of cases.

(Ashley): Okay, thanks for that. I think that would be helpful. I wonder, just to get a sense from the committee on validity, as (Sonny) mentioned, there's a lot of sub-criteria within validity, including the alignment of the specifications, measure intent, validity passing, exclusions, risk adjustment, meaningful differences and disparities.

There's only one data source so we don't really talk about the multiple data sources. But if there's any concerns about any of those before we call this for a vote?

Okay, hearing none, let's go ahead and have everyone vote through the SurveyMonkey. We're voting on validity for Measure 3509.

Woman: Just quick question, do we leave everything else blank for the ones that we did earlier on the other system?

(Ashley): Yes, uh-huh. You can leave them blank.

(Herb Wong): So I have a question. This is (Herb). I thought I heard (Troy) on this call. He was I thought the (pent) person here so I kind of wonder whether or not you actually have a quorum.

(Ashley): We do have quorum, but given the kind of (unintelligible) nature of our quorum, we wanted to just make sure that we don't have to keep clicking back and forth where…

((Crosstalk))

(Ashley): …in the platforms. It just - it caused a little bit of confusion and we're trying to keep it as simple as possible. But we do have quorum. Thank you, (Herb) for asking that one.

(Herb Wong): Okay. Very good. I just want to make sure that was clear before I do the SurveyMonkey. Thank you.

(Ashley): Yes, thank you. You're right.

So we'll go ahead and move on to the feasibility. You know…

(Lina Walker): This is (Lina). Can I just clarify? So only on validity, we don't vote on anything else? Is that - there's not - is there anything else?

(Ashley): I'm sorry, I didn't hear the second part of your question. You said only on validity and then I didn't hear that.

(Lina Walker): Yes, I just want to make sure we don't vote - we don't respond to any of the other questions on the SurveyMonkey. It's just Question 5. Is that correct?

(Ashley): Yes. I don't have it in front of me but there should be one question related to validity for your reading.

(Lina Walker): Okay, okay. Just wanted to be double sure. Thank you.

(Ashley): Yes, we're going to keep going through the other criteria and you can submit your vote as we discuss each of those.

Man: Yes, (Ashley), it's Question 5. So yes, we can move on to the other criteria.

(Ashley): Okay, okay.

Man: Thanks.

(Ashley): Thanks. Okay.

Woman: Okay, yes, and I am here. So with this one we'll go fast. Feasibility as the extent to which the specifications require measure data that are readily

available or could be captured without undue burden and can be implemented for performance measurement.

Given that it's all claims, not fee based, and while specified, there were no comments from the committee. I have this one question for the developer which is while it's all specified in a didn't claim format, if someone - if these providers wanted to reproduce it, do they also get code or you give this back then they get the claims themselves and try to reproduce it?

Man: Thanks for that. We'll have (SASKA) available on-demand that the people can use to reproduce the measures.

Woman: Okay, thank you.

Any committee members have any questions?

Woman: None here.

(Ashley): If folks still like this, this is pretty straightforward and you guys are ready for vote - a vote. We can go ahead and do that. Okay.

Okay.

(Christine Martin-Anderson): So go on to feasibility?

(Ashley): Yes, (Christine). Thank you.

(Christine Martin-Anderson): The first criteria on here is about accountability and transparency and whether or not the performance results are used in at least one (co-

morbidity) application within three years. I think that's been addressed with the purpose of the measures going into the program.

And then the ability to get feedback on a measure from those being measured which has been addressed through the field testing. And then it is not publicly reported but it has a plan used - well is it - will we call it an accountability program? I would say yes. The (unintelligible) incentive payment system.

So the - then there has been extensive feedback. So when we did the poll for use for accountability and transparency, there were some comments about the fact that it's not yet public or it's not quite yet in use and maybe not as well documented. But I do think the developer addressed all of the issues. And I'll stop (unintelligible) to ask you if there's any questions on that first part of the criteria on either feedback, accountability and transparency, and usability.

(Ashley): Thanks, (Christine). This is (Ashley). I just want to clarify that for new measure they do not necessarily have to demonstrate use yet. It's must pass for maintenance measure, but certainly because we know kind of this measure has already (unintelligible) for use in this we have some idea of how it's going to be used. But the bar is not as high for new measures as it is for maintenance measures. I just wanted to clarify that.

(Christine Martin-Anderson): Thanks for the clarification. Okay. Any committee members have any questions for the developers or comments about these?

One measure - I mean, one committee member did comment that there was a test but they needed some discussion around the test. Has that been resolved through this - those hours thus far or is this - we still didn't have that discussion?

Okay, hearing none, we'll go on to usability. From a personal perspective, I think this was a little bit harder demonstrating that in fact there are improvement programs that demonstrate that it can lead to improvement. I think we've had this discussion already around avoiding complications, you know, and not quite so much beyond that on what else the provider could do to affect the measures.

So there's a number of comments that says - some say not well defined, some say not really, some say yes, some say not adequately described. So all over the place on usability. And then again on benefits and harms there were some questions around the specifications around this health cost this patient did some consultations, to discourage paperwork and these around risks or harms the common (leverage) would look at the relative magnitude and discuss how strong the incentive is.

There was another comment about care splinting, (did the lack of latter) adjustment potential risk factors, and then about the benefiting low, you know, relative to the cost of producing the measures and I think the argument also the public comment was the brown's - the pass-through drugs and how that might create a - which we've already discussed, but where it might create a misaligned incentive issue.

So are there any - who wants to discuss the discussion on usability?

(John Wrangler): (Christine), this is (John Wrangler). I think those are really great comments. I just like to echo in terms of the opportunities for improvement. That really was caught me on this measure, I just couldn't wrap my head around.

Because know - I'm practically all from (unintelligible) time to get my mid report from this measure and the radius are so tight, the complication rate is

less than 1% already for the high cost complications that we discussed earlier. I just don't know how this is going to be used to drive improvement and I just couldn't like - I couldn't come up with a good conception for what is it going to do to influence an individual ophthalmologist practice to drive all the improvement with the cataract procedure.

Obviously it's huge impact for my Medicare's frame reference. But from individual practitioners, the radius are so tight. I'm just having trouble with like the improvement of usability.

(Sheryl Jenner): This is (Sheryl). I would concur with his sentiments. I struggled and it maybe that I don't know (unintelligible) sort of what could be done to try to reduce complications but I did not see in the materials presented sort of a tight argument about what say the higher spend physicians could do differently within that episode that would bring down cost closer to average. So I am struggling with the usability (unintelligible) as well.

(Troy Pfizenger): As I know we've had the developer weigh in on this comment at the beginning of our session, but perhaps a brief response conceptually on what the accountable unit can do to influence the outcome would be helpful before I move to the next measure.

(Ashley): Thanks, (Troy), and I agree. Acumen Team, is there anyone that can respond to that?

Man: Absolutely. Unfortunately, (Suzanne) has to hop on a flight so I'm going to kind of I'm going to speak to this point (unintelligible). So there are a couple of things and the points that (unintelligible) up to that about kind of the type interquartile range which came up very early on. It's important.

But one of the kind of the areas to think about this are, yes, complication rates are low but also across a physician doing a number of these. And so there's this thought that, you know, if complication rates are sitting around 2% and your complication rate is 4%, well those extra 2% that you do even though it's spread across a hundred episode is not going to look like that much.

But if you are to (unintelligible) to the mean or to improve and get down to 1% from 4%, that's going to be a relatively large improvement across your - it's going to look small across the 100 episodes, but it is going to basically pan up that you are doing better and cheaper.

On top of this, there are a number of other things such as access offices for cataract surgery, access office-based testing that doesn't need to be done and various other things that in the aggregate, each one is not empirically charged, maybe $40, $50, maybe even more than. But by doing these - doing this excess testing, you are - across a number of offices, you're basically suggesting Medicare to a large cost in aggregate if you're doing the same thing across a number of episodes.

So that is one way to think about this. And then the other to think about, and then the other approach to this and this is based field testing, and what we tried to do there is in a way to really make it feasible for the provider that's only going to be able to take at most an hour to look at this report or maybe 30 minutes is we group all of these down into clinical categories which we showed you analysis regarding these clinical categories but we broke it down and showed how providers did compared to both the national average , as well as provider who treated the same with a docile patient.

And the idea there was to show how you're doing on each of these categories and make something that if you're only going to look at this for 30 minutes,

just very quickly like, "On office-based testing it looks like I'm doing this in 30% more than the people in my same risk docile and on average at about $40 more expensive across all of my episodes. Well that's something that I can look at and address.

In addition to this, within those field test reports, they were provided the list of beneficiaries that they treated, their list of offices. And so some of the GI doctors that actually looked back and looks through those episodes to understand why some of them are more expensive, why some of them were cheaper.

And this was an option for the ophthalmologists that either have the bandwidth or really did want to dig into their reports to look back at their specific patients and understand why some episodes are going to be more expensive and cheap - more expensive or cheaper. And this was something that we did this with the intent that these measures that we don't want to just give people a score, we wanted to give them something to be able to react to in that thought.

(Trey): And this is (Trey) from Acumen. Just to add to that real quickly, after the field testing, the subcommittee had the chance to look at the measure score distributions, (fashionally), ophthalmologists receiving the measures had the chance to respond.

And what we've seen is a reaction where people were happy with the measure specifications. And so any changes that were made after field testing were relatively minor and placed into the measure and then it went into the rule making process as well where there's another opportunity in the public comment.

So I sort of see the measure as, you know, you could have a measure that groups a lot of cost of potentially unrelated things and you could get very high interquartile ranges. Or you could group cost that specialist in this area, the ones that are nominated by the specialty societies, decided we're actually under their influence. And you get a narrow interquartile range but at the same the costs that are in the measure are ones that these nominees from those specialty societies believe are ones they could actually influence and do something about.

And I think that's why the public comment process in rule making went as smoothly as it did for this measure that it's something that people do find to be a useful measure to them as they figure out what their measure set and (mix) is going to look like.

Man: Thank you for those comments from Acumen. So essentially for the committee, you know, we had a lot of robust conversation on this topic at the beginning of the discussion and again now I think we should consider these as inputs on the use and usability question.

And, (Christine), is there anything else new that we would offer or can we consider these inputs, encourage everybody to vote and we can move on to the next measure for consideration?

(Christine Martin-Anderson): My sense is we could vote given that we've talked about mostly the consensus issues have been discussed already.

Man: Okay, thank you for that.

So I encourage the rest of the committees to take these comments under consideration from the developer and (Christine) as well and we can potentially move on, (Ashley), to the next measure.

(Ashley): Thanks. I just want to make our credit…

((Crosstalk))

(Ashley): Yes, there's going to be two votes for use and usability; one for use and one for usability. And then the final vote would be for the recommendation for endorsement. So unless there are any finally comments that you'd like to make before you do your vote for recommendation for endorsement, we can go ahead and do that vote or there's like I said, there's other comments we can hear those now. Otherwise, we can vote and move on to the next measure.

Okay, hearing none, I think it sounds like we are going to have you vote for use, usability and your recommendation for endorsement for 3509 in the SurveyMonkey. And we will queue up our exec measure which is 3510.

Did also just want to do a quick time check and we've got about an hour left. If we could make it through each measure in about 30 minutes, I think we can hopefully end this call having gotten through all the measures and not have to do a follow-up call.

We - there may be points in the discussion, maybe for feasibility for example where you can like your vote for the first measure would be the same for the other two measures and we can just agree that those votes get carried forward. We may able to find some efficiencies there, but otherwise we'll need to kind of go measure by measure and criteria by criteria and have those discussions and then to vote and we'll try to do that in about 30 minutes each.

So we'll do our best to - between (Sheryl) and I to keep everyone on track and (unintelligible) as well.

Man: Okay.

(Herb Wong): (Ashley), this is from (Herb). I'm having a slight problem with the SurveyMonkey. It seems like it had closed out and when I reopened it, I'm on 3510, not 3509.

(Ashley): Okay. Are you thinking that your votes got submitted for the wrong measure? Or…

(Herb Wong): Yes. So - yes, I guess that's my concern.

(Ashley): Okay.

Man: (Ashley), why don't we have the staff work with (Herb) offline and maybe we can have the developer…

(Herb Wong): Sure.

Man: …introduce the measure so we can continue to move forward?

(Ashley): Perfect. Thank you, sir.

Man: Yes.

(Ashley): So Acumen, if you guys can give us just a really brief overview about two to three minutes of the colonoscopy measure?

(Troy Pfizenger): This is (Troy). (Unintelligible) quickly. I must drop off the call for several meetings.

(Ashley): Oh, okay. Thanks, (Troy).

Man: Bye.

(Ashley): (Troy)? Okay, never mind.

Acumen team, if you could go ahead?

(Norman): Absolutely. So this is (Norman) again. Just to really quickly discuss the screening in surveillance colonoscopy. So in kind of difference to the other two measures, the sub-committee that or the committee that made this measure was very focused on the idea that the colonoscopy has a very limited time window of complications and also excluding the ability to include the cost of prep which kind of go across Part B and Part D.

They didn't necessarily think that there were costs prior to the colonoscopy itself that they could really include as (unintelligible) that's prior to the colonoscopy shouldn't so basically included in the global.

So with that they decided to have no pre-trigger window in a 14-day post-trigger window. With extensive discussion about the length of that post-trigger window. The idea there being that they know that within 14 days 95-plus-percent of the complications occur.

And this is something that they discussed at length even thinking about seven days (known) but probably 80% to 85% occur within seven days. And their

thought there was at 14 days they are capturing kind of the expense where complications occur without going too far out where they're going to start capturing noise within trying to - versus capturing complications. And this was kind of the tradeoff there.

And then they also had a very extensive discussion regarding the idea of how long and which to include repeat colonoscopy. And within that discussion, they again did go back and forth about whether to extend this out to 90 or 108 days versus keeping it at 14 days with the idea that there is this possibility that people will begin to push out repeat colonoscopy to report back for other reasons out to 15, 16, 17 days.

And this is something that the choice of 14 days was made for the acknowledgement of that tradeoff and with the idea that this is something that is - they don't really have great data on or more guidelines on. And it's something that they want to keep it at best with the understanding that they're going to continue to look at this in the future in reevaluation.

Additionally, there are measures of quality that look at the number of colonoscopy done within a year, within three years basically to assess colon cancer screening rate.

So without going too much further, I wanted to bring up a couple of points. First topology which I discussed in the first hour (unintelligible) discussion that in both the tradeoff of capturing those people that are biopsying and sending out way too much topology versus the people that are not - that maybe seeing - better at identifying cancer effusion.

The second is in the use of anesthesia and this is something that was also discussed at length. Anesthesia can be overused and start - anesthesia sometimes also appropriately used versus conscious sedation.

So what I really want to get around is similarly to the cataract measure, all of these decisions were discussed at length especially given kind of the very specific group that they focus in on ultimately screening surveillance colonoscopy excluding those patients that had IDD. These are higher polyp risk, excluding any diagnostic colonoscopies and very much focusing on homogenous patient cohort.

It was very easily able to focus in on all of these difficult decisions that have been not only going on in the GI community but because of that really came to the forefront and try - and deciding on the specifications to this measure.

So with that, I will let you all go for it and discuss.

(Ashley): Thank you. So I think we're back with (Lina). Are you there?

(Lina Walker): Yes, I'm here. And I'll be pretty quick. So the - well I think that the developer covered the - much of what I was going to say in terms of what they had presented in the document. The, you know, screening colonoscopy has become the most common screening test. They do say that in the opportunity for improvement that it was (on recommendation) and I think the developer mentioned use of anesthesia can be and was one example of a big use and also they noticed in the document inadequate bowel preparation and then cited some studies related to that.

For the committee comments, it was similar to the cataract comment in the sense that the majority thought that there was high impact to high resource use

and that that would be opportunity for improvement. But the committee members who were very concerned about the very small (interquartile) range. The IQR is only in the range of $170 at the clinician group practice level and at the practitioner level. The means score itself is very small for - at both levels. It's in the $900 range, $936 and $979 specifically.

And I think the two questions that I would like the developer to respond to came from the committee but I also have those questions is, if they can talk a little - say a little bit more about the - show to the - describe the relationship between that variation. It's not a lot of variation but where there is variation, can they - what is the evidence that poor quality is leading to that - driving that variation?

And then the other thing was that there was very little discussion about what the - and so it's a similar question but there's very little discussion about the opportunity to reduce to - that variation within the episodes. So if they could say more about what could be done, for instance, to reduce that variation.

And then to be a little bit more specific about the utilization, I know you mentioned a couple of examples. If you could say a little bit more about that, that will be helpful.

(Norman): Sure. This is (Norman) again. So to first discuss the overutilization, there are actually a couple of things. The first is of course anesthesia. In some cases, anesthesia is definitely wanted and this is something that was discussed by the subcommittee with the specific inclusions of variables that would suggest requirement to necessitate use of anesthesia.

The second is I discussed when I went over pathology and the idea of over biopsying but not only that. There's also a known practice of taking each

biopsy assessment and putting it in separate pathology vials and basically handing over five pathology specimens which each pathology specimen is going to be billed separately as opposed to putting all five in one. And it is - and they expressed how it's known that this doesn't necessarily increase the detection rate of cancer but it does increase the cost of pathology.

So those are two very simple kind of utilization approach that can be used to decrease cost.

The second is with poor bowel prep and also there's something else that we didn't quite get into but basically education. And so with poor bowel prep and with education as a patient, this actually should the - first off, if the colonoscopy can't be done then the colonoscopy needs to be repeated, it escalates the cost. But, say, with poor bowel prep the colonoscopy was done, with poor bowel prep, this necessitates longer intubation time, I meant longer colonoscopy time, as well as kind of - which can lead to higher risk of cardiopulmonary complications within the first three or seven days.

Similarly in educating the patient, there is - there are things that a patient should know after getting a colonoscopy about whether they have pain, diarrhea, things of this nature and that is something that will aggregate some of the ER visits or observation stage for these expected symptoms.

And then finally, the last thing -- and this is a much rare outcome and it was acknowledged as such -- basically after doing a polypectomy or even with the colonoscopy itself, there is always a higher risk of post-polypectomy, post-colonoscopy bleeding.

And while the occurrence is rare, this is definitely a complication and an understood complication and it's something that to some extent is going to

happen and there will always be a percentage of patients who have this. However, there is also - there's also a requirement for the colonoscopist to be aware that a patient is on work on blood thinners and to make sure appropriately to tell the patient when to stop, when to hold or whether they should be held at all.

So those are multiple different ways in which a colonoscopist could directly affect their score.

And specifically similar to cataracts, we're looking at people that are doing a large number of these. And so it does come down to kind of a - that rate problem. So if you are - if you're having cardiopulmonary complications in 5% of your patients and then drop that down to 1% across your 100 episodes, that may not look like an incredibly large amount but it is going to be an amount that should - going to show you being effective in improving your care for your patients.

Man: And just to give you a sense of, like, the scope of the types of improvements that are possible on the measure through either types of channels that (Norman) just talked through, if you look at mean episode cost for the first quintile of episodes, it's on the order of $700, $705 actually. And then if you look at the 5th quintile, it's $1350. And so it's almost a doubling in the spread between the bottom 20% of episodes and the top 20%. And so I think that, in combination with the extraordinarily high reliability numbers here, suggests that for any given episode, there is substantial room for variation in improving or reducing cost.

And the - that low-cost episodes and high-cost episodes seemed to be concentrated among particular clinicians like a high reliability suggests that there was in variance of these costs are relatively low and that despite the

interquartile range that's been brought up, that's a measure they were able to distinguish between clinicians who were able on any given episode to take advantage of those opportunities versus clinicians who are…

Man: Those are great comments. Thank you. Thank you from the developers. So I think these topics are bleeding a little bit into reliability and validity. So we just wanted to check in to make sure, you know, is there anything - it sounds like there's a lot of - in terms of the opportunity for improvement, even if those are interquartile range questions, you know, has been discussed or is a range of performance.

So I just want to see if there's any other comments and then maybe we can ask (Herb) to sort of look in - or provide some discussion on the reliability statistics given that the methods are similar to the method used on the colonoscopy measure or on the prior measure, just at least a brief summary there and then we can dive into the validity questions which has been the majority of the conversation so far.

Any other comments…

Woman: Well do you want us to go ahead and vote on the first…

Man: Yes. Yes please. And then, (Herb), can you get us started on reliability?

(Herb): Yes absolutely. So as, you know, with all of these measures, reliability testing is pretty much similar across all three metrics. Again, the acumen uses the test/retest with correlation signal-to-noise ratios. In terms of looking at the results and comparing it with 30509, they are very similar. The test/retest results for the piercing correlations are 0.93 for the group level and 0.88 for

the individual levels. When they take a look at the split sample, again it compares pretty much similarly with 3509 in terms of - on that component.

When we take a look at, I guess, the signal-to-noise ratio component of it, again, it scores relatively high with 0.96 for the clinician group and 0.93 for just the clinicians. So again, very high.

In terms of staff preliminary reliability rating, the staff rated it "High" in terms of the committee, the individual panel committee that rated it. It did a little bit better than 30509 with four "High" and two "Moderate" votes there.

Comments from members basically saw very little concerns across the board in terms of specification and real liability testing. The general comments focused more a little bit on more details about the different components there. So, you know, looking at 305 - comparing this with 30509, it performs, in terms of reliability, in the similar fashion, maybe a tad better.

Man: Thanks, (Herb). Comments from committee? It seems fairly straightforward related - I mean, the methodology is very similar and, (Herb), a nice job of comparing the performance data.

(Ashley): Hi. This is (Ashley). I just want to clarify again the vote here would be, again, remember the Methods Panel have already voted. And so the committee would just need to decide whether or not to accept the Methods Panel recommendation or if you would like to reach adjudicated (repo).

Man: And, (Ashley), what was the Methods Panel recommendation?

(Ashley): The vote was four "High," two "Moderate." So it was "High."

(Sheryl): Yes. This is (Sheryl). I would concur with the panel's recommendation on reliability.

Man: Thanks, (Sheryl). Other comments in the group or are we good on this subcriteria?

(Ashley): I'm good.

Woman: Good to me.

Man: Okay.

Man: Agree.

Man: All right. Thank you all. So on the validity question, (Sonny), if you can walk us through this component. Again, a lot of these elements where we've discussed already in the preface of this conversation related to the actions that can be taken by the accountable unit to be able to influence the results on the cost performance, the developer walked us through the, you know, improvements to - for better bowel prep education and bleeding complications as components here and the methodology used for the validity testing was very similar in terms of the component of correlation analysis. So, (Jack), if there's anything else new that you want to offer as it relates to the methodology used as a result, we would welcome that. But, (Sonny), you want to kick it off for us?

(Sonny Vilnany): Sure I agree. Pretty much similar methodologies. There are certain differences between the previous measure that we looked about significant differences include the R-squared is here, overall 0.12. There were two members from the scientific method panel which had some concerns about

that and I think we should just talk about that just so that no committee member here has any questions regarding that.

Two other themes, which were clinical relevant with this measure, were pathology cost and bowel prep. I think we talked about that already.

Another issue which was raised was socioeconomic assessment and a person had concerns about there being no conceptual model for this measure and the other measure and the measure - and the developer's alignment on the sole paper from (Pope) which does not take in account SEC but only (ACC).

So I think those are the main high-level points which I think we should talk about or at least take a moment to talk about. Apart from that, I have no other person in point.

Man: (Jack), do you want to kick us off with that on the R-squared if there's anything there and then the component analysis methodology for validity testing?

(Jack Needleman): No, some of these risk adjustments have high R-squared somehow below and I think it reflects how well the, you know, how standardized treatment isn't how much across the different risk adjustment categories and how, you know, how much variability there is across the patients by any of the things we measure. I don't have any concern there.

My concern is still with the pathology cost inclusion. So - and that's personal as opposed to my sort of representing the - summarizing the general discussion of validity. So I'm happy to either kick that off or wait until the general discussion is done.

(Sheryl Jenner): (Jack), can you remind me what the problem is with the pathology? This is (Sheryl).

(Jack Needleman): Well the - they've included pathology cost and some of the folks on the scientific - the external review has questioned the appropriateness of this given the risk for whatever incentive it creates for doing too little biopsy and too little - it's not a case that any pathology is wrong. So the question is whether - and there's some concern the risk adjustment is just not going to capture appropriate variations in pathology because it has to do with the number of polyps that are identified.

So the developers have said there's a concern about separating the samples, so you wind up with high pathology cost for that. But - and I'd like to know how - what information they have about how frequent that is. That would be an abuse that one would worry about.

But the other side of that is they also said in the first presentation that they identified - they looked at the correlation between the adenoma identification rates and the pathology cost and I'd like to know more about what that correlation is and how strong it is as a way of thinking about how much I should worry about the inclusion of - about the potential effect of discouraging pathology and therefore discouraging identification of cancers or pre-cancers.

(Sheryl Jenner): Thank you.

Man: Do other committee members have some feedback or comments related to this topic in particular?

We've discussed the social risk factor question as well. (Sheryl), you know, I don't know that there's anything new necessarily that would be of slide here but I'm…

(Sheryl Jenner): No.

Man: Okay it feels like…

(Sheryl Jenner): Yes I think it's covered.

Woman: Yes.

Man: Yes I agree. Okay. So then it's really - other committee members on the pathology question because that appears to be the main validity discussion for this particular measure. From the committee first.

(Lisa Lock): This is (Lisa). It seems a little bit like you're damned if you do and you're damned if you don't because, you know, I understand the developers' point that we want to discourage bad behavior of oversampling or separately sending things to pathology that he mentioned this morning. So, you know, I mean, I tend to be a fan of the measure and then retest to look for unintended consequences but…

Man: Yes it seems like there just may be two approaches to doing this and we've heard the rationale on both sides of it. Are there other comments from the committee?

From the developer, I know you've had the opportunity to already provide a rationale for inclusion. Anything new to add here briefly?

(Jack Needleman): And specifically - this is (Jack). Specifically I'd like to know about the correlation of the identification rate and pathology costs and what evidence you have that the high pathology costs are being driven by practices that are not associated with better identification.

(Norman): Sure. So hey this is (Norman) again. So a couple of things to add. The first is that one of the things that actually came up in this discussion was the missed adenoma detection rate measure and specifically how one of the things that they wanted to do was to, in effect, align the patient population by not including patients with IBD, not including patients with previous colon cancer, things of that nature with the idea that you are getting at a lot of that - a large amount of that population that would already have a higher risk of having polyps. And so the population that you're left with is hopefully going to be much more homogenous and going to spread that out. So what you're going to actually see is basically people that are biopsying more or sending off more pathology samples.

The other thing - and this is something that we did very quickly and provided to the committee but unfortunately he was done a while back. And so we can try and work at it. But we did provide them with a correlation of really the kind of overall distribution of number of units on pathology billed after colonoscopy, specifically for polyp pathology.

And this actually led off on a series of discussions including whether to only include a certain amount of pathology after X number of bottles because from just kind of my vague recollection of it, what it did show was that there were a number of people, a not insignificant number of people that we're billing for - that were - that for with - are not insignificant number of colonoscopy. There were four, five, six, seven pathology units billed.

Oh, also one other thing I wanted to add and I don't remember if I had mentioned this earlier. Like, after the discussion, even with all of the back and forth, they did end up voting specifically on this question of whether to include or not with the acknowledgement to (Jack)'s concern and overall there was a vote, a 64% to include versus 36% to not include.

Man: Very helpful. Thank you. (Jack), are there any other comments related to this or do folks feel like we've sort of addressed the validity question? One of the main validity questions that's specific to this measure.

(Jack Needleman): This is (Jack). I'd like more data but I'm ready to vote.

Man: Okay. Thank you. Other comments from the committee?

Woman: No. I'm ready to vote.

Man: Okay. So while folks are voting, (Christine), I know that, you know, the use and usability and feasibility a lot of these topics are similar across the measure. So we can certainly, you know, I would encourage if there's specific comments that relate to this measure in particular. Maybe you can highlight those as we get into that next criteria.

(Christine Martin-Anderson): Thank you. There were no unique comments for feasibility that was almost identical. For use, I think the same is - pretty much identical in terms of the same programs been planned and the usability - there was a question about whether or not those measures have seen performance reports. I think that the developer has addressed that and they have through the field test.

There was one particular issue on use around the clinical test and not taking action on those. I think we just covered some of that as well. But I want to see if there's anyone on the call particularly this comment about phlebectomy, et cetera, if we feel like we need to discuss that again around use. And for usability and improvement, I think we've heard a little bit more of a discussion about those things that the developer could do but there are still some committee members who are skeptical of the usability.

So let me just see if anyone wants to discuss any of those before we get to harm.

So there's - so can I take the silence as the discussion so far has at least addressed those issues?

Woman: Yes.

Man: Yes.

(Christine Martin-Anderson): Okay. And then there was in terms of - oh go ahead.

Woman: No I just said yes.

(Christine Martin-Anderson): Okay. On the benefits and harm, I think the one part that there is a comment about (unintelligible) which I haven't heard discussed. So if there's anyone who wants to particularly address this issue around social risk factors to put off screenings and leading to excess costs down the road.

And hearing nothing I'm assuming people are ready to vote or anyone - going once, going twice.

Woman: Ready to vote.

Woman: Ready to vote.

(Christine Martin-Anderson): Okay.

Man: Thanks, (Christine). So while folks are getting ready there, are there any other comments on any of the criteria or the overall measure before, you know, folks are ready to vote on the overall measure? If not, I'd ask the developer to tee up an introduction on the next measure for discussion.

(Rose Do): Thank you. Hi everyone. This is (Rose Do). I think if everybody is ready, I can talk about knee arthroplasty.

Man: Please do. We can limit the comments to about two to three minutes. It'd be ideal so that the committee has some time to review the measure.

(Rose Do): That sounds good. And I'll also try to focus on the specific critiques that were raised.

So just wanted to introduce myself, I'm (Rose Do). I'm one of the physicians and a medical officer at Acumen. Pleasure to be on. I hear some familiar voices on the line and I'm going to give a brief overview of the knee arthroplasty measures. And you also had heard that one of our co-chairs, Dr. (AJ Yates), in orthopedic surgery called in as well. Our other co-chair, Juan Quintana, is currently in a surgical case. So unfortunately he couldn't join but I wanted to thank Dr. (Yates) and all of you for making the time today.

So we discussed the process earlier and I'll just kind of talk about the methodology specific to this measure and then highlight our clarifications for some of the items that were brought up on the worksheet.

Regarding the process, our co-chairs led a group of subject matter experts in the field, such as orthopedic surgery, physical therapy, geriatrics, general medicine, anesthesiology and many others. Some of our experts have sort of surgical committees for NQF. They've been involved in the bundled program and hospital measures for PKA in hip. So such there were efforts to harmonize with that work.

Notably this measure that we're talking about is in sync with NQF 1550 for a hospital level standardized complication rates in hip and knee. So there's similarities in the specification with what's currently endorsed.

So the window for the cost calculation is the preoperative planning period in the 30 days prior to knee arthroplasty. And then it extends to 90 days post-op. With all the subsequent measure specifications, this decision, as well as the others, were guided by careful analysis, demonstrating the detection of important complications within 90 days. The episode is triggered by two CPT codes for a knee joint repair and then the workgroup members also felt strongly that we should narrow our patient cohort to exclude the surgical procedures for infection.

We use modifier codes to treat subgroup for unilateral and bilateral surgeries within the episode window. And the attributed clinician was the orthopedic surgeon that was performing and doing the TKA.

So cost components or the services assigned to the measure were selected to only include preoperative imaging, lab work, consultation, date-specific

physical therapy. For the postoperative period, the work with members helped us to find important complications such as EDP, pulmonary embolus, wound infection, bleeding and others. And then again many of these outcomes are modeled for the NQF 1550 so that we could have some alignment.

So I just want to highlight a few areas that were brought up in the comments. So we talked a little bit about post-acute care cost. During the eight-month process of measure construction, our subject matter experts discussed the implications of including those costs. There's been a lot of attention to post-acute care due to the high cost, the Institute of Medicine report, the impact as well as many other analyses and we know that in bundled payments, providers have to make choices on when and what type of post-acute care to use.

We have just in policy that considers post-acute care. So for instance, the hospital-based measure NQF 3474 for hip and knee arthroplasty includes payment for post-acute care. With that in mind, the work with members included path of post-acute care but they decided to more granularly define when to include those costs. So for instance, we only use skilled nursing facility with qualifying hospitalization. We use diagnostic and timing information and we apply those throughout the entire episode window.

And when we talk about this, our experts felt that the risk adjusters, the time windows, grouping rules, they provided adequate protection against any consequences for this and they felt that, you know, despite the trade-offs that we're going to be discussing, post-acute care should be measured to understand the (unintelligible) that's used, incentivized coordination of care between the hospitals and post-acute care facilities, and then also the treating and discharging clinicians.

So we've also talked about, you know, aligning the cost with quality and we want to make note that there is a patient reported outcome measure, NQF 2653, on functional status after a knee arthroplasty. So I just wanted to keep that in mind with this cost measure.

We talked about the social risk factors. (Alice) and (Shree) discussed this in the beginning of the call. I'm happy to discuss again if we need to. But I just - I think I wanted to add that we have done the internal analysis from the current hip and knee measures. And then some of the other new measures do not include more status. So we've done it on analysis and talked about those.

Man: Thank you very much.

(Rose Do): And then finally - oh it's okay.

Man: Okay go ahead. No sorry. Please finish.

(Rose Do): Oh no sorry. I am almost done. Sorry. So I just wanted to mention the risk adjustment methodologies that we used the (ATC). But then the orthopedic surgeons gave a few other things to include that were specific to the measure. So knee flexion contracture, opioid use history, psoriatic arthritis, very important measure-specific comorbidities.

At the end of the process, you know, we talked with a number of practicing surgeons who felt like this measure does harmonize well with existing measures and it accurately reflects real-world clinical care and we also include these adjustments that are not seen in any of the other endorsed measures and included related costs that are at a very granular level.

So I can pause right there and we'll take any questions.

Man: Thank you for those comments. Very helpful. With that, (Lina), can we get started with the importance to measure and report and we'll address questions as we go to each of the criteria? (Lina), any specific comments here related to opportunity for improvement?

(Lina Walker): Very quick, so all the committee members who voted expressed that this is high-volume significant spending. It was high (resource) use. I felt that that was opportunity for improvement and that the developer had done (unintelligible) point.

Man: Thank you, (Lina). Comments from the committee in importance?

If not, I would encourage you to work through the SurveyMonkey for your voting. And I'd also tee up (Herb) for the reliability. But any other comments related to importance to measure?

(Herb): Okay, so I'll get started on the reliability. So again, Acumen basically applies the same reliability testing. I'll cut right to the chase in terms of the test/retest results. And their peers in correlation is 0.8 for the group level and 0.75 at the clinical level. This compares somewhat lower than the other two measures that we looked at before. Just as a reminder, that range was between 0.88 versus 0.93 but still in a very good range there.

In terms of the signal-to-noise analysis, the mean reliability score was 0.87 for the group and 0.81 for the clinician. And again, in terms of comparison with the other two measures, that ranged from 0.93 to 0.96, again, within a good range.

In terms of the staff recommendation or their general assessment, they rated it as a "Moderate." This is reflected in the Methods Panel individual reliability rating that against had one for "High," four for "Moderate" and one for "Low." So the overall assessment in general was a "Moderate" on this. So of the three, this one probably performed just a tad lower than the other two.

In terms of questions, the questions and comments were similar to the other two. Most folks did not have any comments, per se. Others have comments that - or seeking clarification or looking for additional details in terms of (unintelligible).

Man: Thank you, (Herb). Comments from the committee? Methodology is very similar. Really the question is weighing the output.

(Jack Needleman): Yes. This is (Jack). I found the presentation of reliability in the, you know, initial presentation when I looked at it from the Scientific Methods Committee is right on the margin. Actually for me it was a little bit below the margin. I was in the minority there. And one of the things I really appreciate that Acumen did is they went back and provided some additional analysis, including given - as they know, when you'd split the sample, the standard areas get bigger because you've got half the sample. So they did some reanalysis of the quintiles movement based upon - well, they actually didn't describe the methodology. And I'm not going to ask them to do it on the phone call. But at some point, the next time you do it I'd like to see the methodology better described. But I assume it was looking at the standard errors around each physician and making some prediction about the likelihood that they would move on a second sampling. And the numbers look a little better on that on the reliability side when you do it that way.

Some of the signal-to-noise for some of the smaller for - at the low end are a little bit low, again right on the margin. And I the question I think the committee had was whether the number of cases ought to be pushed up to just make the reliability measures at the low end for low-volume physicians a little bit tighter. I do think that there's an argument to be made that this can be comfortably applied at the MPI - at the (TIN) level but not necessarily at the MPI level.

Man: Right. Thanks for those comments, (Jack).

Other comments related to this? And I would ask (Sonny) to be in the queue in terms of the validity conversation. But any other comments related to the reliability performance?

((Crosstalk))

(Ashley): This is (Ashley). I just want to make a comment here. If - you know, again the vote here for reliability is whether or not to accept the Methods Panel vote. And so, you know, if there are other concerns, I just want to make sure that there's an opportunity to discuss whether or not you'd like to do that or if you'd like to submit your own vote.

Man: (Ashley), what was the results again? Sorry to keep putting you on the spot, just so everybody is aware of what they're agreeing to.

(Ashley): Right. I think (Herb) summarized it for us but basically the Methods Panel ratings were one "High," four "Moderate," one "Low." So based in kind of averaging all those together, the rating ends up being a "Moderate."

Man: Yes. Comments? Hearing no - I mean, hearing - if I don't hear comments, I'm assuming that's in support of where the Methods Panel landed. And in some ways, (Jack) also pointed out - again, credit to the developers here that provided some additional analyses but essentially they - you know, was it a little bit more information provided? That's maybe even a little bit better reliability. But thoughts…

((Crosstalk))

(Sheryl Jenner): Yes. So just to…

((Crosstalk))

Man: Go ahead.

(Sheryl Jenner): Yes. Yes. This is (Sheryl). You know, in the context of the comments (Jack) made, I actually think this falls into the higher range. So I would want to be able to rate this one "High." Because the panel's rating was "Moderate," correct?

Man: Yes. Okay, so I - in the way that the survey is presented, I would encourage you then to vote in the way that you would based on the information that was presented by the developers and the conversation here.

Woman: So we have Question 19 and Question 20. So you can skip Question 19 which is do you accept the Scientific Methods Panel range for reliability and move right onto Question 20. And that's where you can fill out the ranking and the rating for the reliability.

Man: Is that clear?

Woman: Yes.

Woman: Yes. Or you can answer it yes, you know, no and then - right? There should be a yes/no to that, (Kate)?

(Kate): Yes. You're right.

Woman: Okay. If I can - you can answer No to 19 and then submit your vote for 20.

Man: Okay. So I think certainly there's been a lot of conversation on the reliability element. (Sonny), can we ask you to sort of kick off the validity conversation, again pointing out that the empirical validity testing comments and approach has been very similar, pointing out specifically resource utilization post-acute care period and then also the overall R-squared statistics were provided.

But any other specific comments that you'd like to raise related to validity?

(Sonny Vilnany): Yes. So there are a few comments that I wanted to raise about validity. First of all, a few people aptly pointed out that (unintelligible) associated with your status had a huge range from 467 to 3586 which were meaningful numbers. And I think since several of the people on this committee and SMP members have mentioned that we should spend a time talking about this to see if there're any concerns about that, there were some concerns about post-acute services. I don't think - I think the developer did a good job addressing them. I don't have any major concerns about that.

And that was pretty much the only thing that came to my mind. I think the others have been addressed in prior measures and prior conversations here.

Man: So, (Sonny), that was specifically related to the SDS, the impact of SDS?

(Sonny Vilnany): Correct, the dual status

Man: Right. Okay. So can we open that up to discussion? Do folks have thoughts that have - you know, there's certainly been a lot of conversation and materials. Does the committee have or the committee members who have some thoughts related to this question?

And, (Jack), maybe also - oh yes this wasn't considered in the Scientific Methods Panel but - okay, comments from the committee?

(Jack Needleman): This is (Jack Needleman). I am more concerned about the post-acute cost and the duals than has been expressed. And I'm willing to - not withstanding that it's a smaller square, there was $1000 difference in treatments in duals and the non-duals, mean difference. And that's pretty substantial. And I would have a preference for seeing the measure stratified for the duals.

But that to me - that should - I'd be open to seeing that reconsidered on the revision, you know, down the road. I am more concerned about the post-acute. And the additions that acumen made -- and unfortunately, I don't have the document in front of me because I'm out of my office -- in the subsequent addition that they made to the document using their clinical themes made that even more substantial to me.

There were a number of areas where they looked at the correlation by clinical theme of the costs to - specific cost within it. They noted that the correlation with complications was high and that's quality issue. That with one of the other cost measures -- and again I apologize for not having it in front of me -- was low. And that was appropriate and comforting. But they said that the

association correlation with post-acute was the highest. And in their write-up showing that it was valid, they mentioned the high correlation with complications. They mentioned the low correlation with the other measurement. They ignored the high correlation with post-acute.

I'm not completely convinced because we haven't seen any analysis that the risk adjustment is fully adjusting for appropriate differences in post-acute. I am concerned about under service, the incentive for under service on post-acute. There's going to be a lot of variability on which patients need post-acute and how much they need and that appropriate variation ought not be penalized than a cost measure. And I'm not convinced that this measure as constructed has done it. There were other options including accepting a minimum level of post-acute, maybe going up to the median, maybe going up to the 75th percentile and looking for cost above that.

That was not the path that was chosen here. And I - given that high correlation of this overall cost with variations and the overall cost with variation in post-acute and the lack of clinical discussion of whether it can be clearly identified which patients are in need - going to be in need of more post-acute care and that the risk adjuster has effectively control for that, I don't think the inclusion of these costs are valid if it's a measure of - if we're trying to capture inappropriate high and inappropriately low-cost efficiency rather than differences in patient need.

Man: Thanks, (Jack). So before we get into conversation about this, I just want to query the committee. There seems to be a lot of - or there seems to be two main issues that the committee will have to weigh in their voting of this measure. The first is related to the handling of duals, so not including the duals considering the magnitude of the differences and then (Jack)'s last point here on the post-acute care services.

So before we move on to discussion of that, are there other issues that the committee would like to discuss related to validity?

Okay, hearing none. On these two issues in particular, do the committee - do other committee members have comments on this? Or maybe we can open it up for the developer for a quick response on these two issues for consideration for the committee. But ultimately this is the methodology that's in front of you and the committee will have to make that determination in the voting. And obviously I would queue up (Christine) for the next criteria. But do developers have response on these two issues, in particular on the methodology decisions that were used for both these issues?

(Ashley): And this is (Ashley).

((Crosstalk))

(Ashley): I would ask that you keep it very brief to like one or two minutes please.

(Rose Do): So in the interest of time, I'd like to ask if (AJ), our co-chair, had some comments because I know that he wanted to share.

Dr. (AJ Yates): Yes. I can understand the concerns about the post-acute cost being variable. But that's actually where the greatest variability is going to be. And it's - there's a lot of surgeon decision making that goes into that.

I'm in the middle of a bundle being in the CJR bundle here and I can tell you that one of our local hospitals within our system had 60% of their patients going to post-acute care skilled nursing facilities as a habit. They subsequently dropped that down to 15%. This is true across the country that

anybody within the CJR or other bundled plans defines that the their greatest savings in terms of probably inappropriate spending comes from (prehab) and from preparing the patient with education and making sure that you've developed a plan to get the patient safely to home and they stay home. There's no evidence that going to a skilled nursing facility improves the quality of outcome. There's no evidence that IRS improves the quality of outcome.

And so if you're going to look at the cost, you have to include post-acute care as a free floating source of variation. And that source of variation is rapidly changing to becoming much more narrow. I suspect that some of the dual eligibility questions because of socioeconomic reasons might mirror some of that or might be contributing to that change as well. This is not dissimilar to what's in the hospital cost measure which is mentioned at the beginning of this. And you - because your committee is already endorsed, I was on the TEP for that as well. Thank you.

Man: Thank you for those comments. Other committee members on these issues? Clearly, you know, these are decisions that were made. Other standing committee members have any other feedback on these issues?

Okay. (Christine)…

((Crosstalk))

(Lina Walker): This is (Lina). This is (Lina). I just like to react to the developer's comment. So I guess, I mean, hearing that makes me worried about these unintended consequences of sending patients home too soon without appropriate post-operative care. And, I mean, I'm hearing that there's no evidence that going to a nursing home or skilled nursing facility improves outcome. But like

what's the evidence that sending patients home sooner is better for the patients as well?

So, I mean, I wasn't worried before but now I am worried about including these post-acute costs because I don't feel like that explanation was very compelling and in fact makes me more worried now about unintended consequences.

(Jack Needleman): Just - I have to reply to that since I've been called on. I think those are the rules for the debate tonight and last night. But the fact is is that we're not acting in a vacuum. We're very much under the radar, on spot in terms of quality metrics. And so none of these happens without there being guarantees in terms of quality. And I can tell you that a decrease in the utilization of skilled nursing facilities has not been accompanied by any sign of increased complications, increased readmissions, increased dissatisfaction. And if anything, satisfaction goes up.

So there are checks and balances to this in terms of this being implemented across the country. The checks and balances just factorily hold on to the fact or demonstrate that the perceived possibility of unintended consequences hasn't happened. And that's coming from a facility that's had to respond very - you know, we're one of the largest medical centers in the East Coast and we do not see any change in our qualities - quality metrics. And that hasn't been seen by Medicare and it hasn't been seen as an unintended consequence.

I write about unintended consequences and worry about it all the time especially for patients that are on the border of being a higher risk. And I worry about condition classes having - being hurt by people trying to avoid them. And that's something that I worry about all the time. That's the biggest

thing that you have to worry about with any of these under the microscope sort of things.

But in terms of the outcomes, people going home have been just as safe as they were before. And I - unfortunately I can't present that to you in detail because this is supposed to be a short comment. But I could come up with multiple papers showing that that just hasn't been born to be true.

Man: Thank you for those…

((Crosstalk))

(Janice Orlowski): This is (Janice Orlowski). In the (AAMC)'s experience with bundled payments for CJR is exactly as the previous speaker. We have found that with appropriate preparation that there is no difference as we take a look at the use of skilled nursing homes, SNPs or rehab. And that high quality has been maintained. And that actually was the basis for CMS making the decision to have the CJR program be a national program because they saw the improvement in cost with as good or better quality outcomes with the preparation.

So again, our experience mirrors the previous comments.

Man: Thank you, (Janice). And I think we're also getting a little into usability and use based on the unintended consequences conversation and harms.

(Christine), can I ask you to kick off that conversation in terms of what maybe we should be considering in terms of feasibility, usability and use as a committee?

(Christine Martin-Anderson): Yes. Sure.

Man: Go ahead.

(Christine Martin-Anderson): Feasibility is the same as the other two measures. Identical comments. Use, the same. Usability again is the bigger one, focused on is the measure driving improvement and whether it will improve cost or efficiency. But in this one a little bit more of a dialogue about the relationship between cost and quality and maybe there being an assumption about the direction that goes and not being sure.

More - I just wanted more discussion on avoidable cost. And then I think on benefits versus harm, there was - there were comments about whether or not it will discourage appropriate post-acute services, which we just discussed the care (unintelligible) comments again. And that's pretty much it. Transparency again not an issue.

Man: Thanks, (Christine). And again, I know we're getting towards the end of the time here. But I want to make sure that the committees had an opportunity to discuss the benefits and harm's question which very much relates to the question of validity on post-acute care service and duals.

Are there any other committee comments on those two topics? Clearly there's a lot of discussion in the material. This has been discussed quite a bit in the SMP. It's in the preliminary analysis. So any other comments that the committee would like to share related to those items in particular?

Any other use and usability or feasibility comments?

Any other comments on the overall measure before we move to vote on the overall measure?

So I think it's time for us to open it up for public comments to be able to make sure that the folks who've been listening in intently on this conversation have an opportunity to weigh in. Could we open up the lines? Are there any public comments from any members of the public on the measures and the consideration of the measures in front of the committee?

Are there any comments on the Web chat?

(Ashley): No. Not that I can see. This is (Ashley).

Man: Are there any other comments that the developer would like to make related to the discussion? I know we kind of moved through these items expeditiously. Just before we close out.

(Rose Do): Sure, really quick. This is Rose. I just did want to reiterate that we do attempt to align with some of the quality measures that are already in existence. I mentioned some of the patient-reported outcome measures. We have readmission measures as well. And given that this is a cost measure, any of the things that folks are doing, clinicians are doing to try to save cost from a post-acute care side can be captured in terms of cost as well. So readmissions will be factored into cost measures. And then in the greater landscape I just wanted to remind folks about our overall alignment is the quality.

Regarding post-acute care and - or sorry, the dual eligibles, we also do our internal analyses and notice that there wasn't much change in terms of scoring for providers. So we do see differences in (unintelligible). But in terms of

changing the mix score, it's actually a very small number of clinicians who experience that.

Man: Thank you for those comments.

(Kate): Okay. So this is (Kate). And we will quickly go on to our next steps. We'll be in contact with the committee regarding our next meeting. We will have follow-up communications. There were a couple of things that we didn't get to discuss today. So we'll be able to communicate those either via e-mail or via call. We'll make a determination.

And our draft report is going to be out for a 30-day public commenting, July 5th to July 25th. And then we will have a committee post-comment Web meeting on September 25th from 2:00 to 4:00 pm Eastern.

And since we are at time, we are just going to move on to the contact information. If there're any questions or concerns, please always feel free to contact us at efficiency@qualityform.org. The phone number is there as well as the Project page and SharePoint information.

And with that, we thank you all so much for joining us.

Man: And, (Kate), on behalf of the NQF team, I just want to thank the co-chairs, in particular (Sheryl) and (Brett). And I know (Brett) may be stepping down as co-chair. We want to appreciate the several years of hard work in particular for the co-chairs that have been critical in this work and then obviously to all the members of the standing committee for your comprehensive review and the lead discussions in particular. You guys all make this. There's obviously a lot of work that is in place here but obviously very important for our - the measurement community.

But, (Sheryl) and (Brett), any other closing comments from you?

(Sheryl Jenner): No. I just want to thank the measure developers for all their good inputs and (unintelligible). There was a lot of documentation to go through trying to extract development pieces. It's always a challenging endeavor. But thank you so much.

Man: Okay. All right. Thank you all for your time.

Woman: Thanks everyone. We'll be in touch. Bye.

Woman: Bye.

Woman: Bye-bye.

END