NATIONAL QUALITY FORUM

Moderator: Ashlie Wilbon
June 22, 2011
11:00 am CT

Operator: Good day and welcome to the Resource Use Steering Committee conference. As a reminder, today's call is being recorded.

At this time for opening remarks and introductions, I would like to turn the call over to your host, Miss Ashlie Wilbon. Please go ahead, ma'am.

Ashlie Wilbon: Okay. Thank you. Welcome everyone. Thanks for joining us this afternoon. This call is a call for the Resource Use Steering Committee. The Ingenix measure 1599 and we're going to be talking - focusing on the scientific acceptability of that measure. And just a brief review, I'm not going to do a roll call this time. We can get that actually through the conference centers. We can see who dialed in.

But I did want to let everyone know who's on the call that Doris Lotz who was the co-chair up until - for the last year - so had to step down from the committee she had too many conflicting obligations and didn't really have the time that she wanted to devote to the project.

But on a good note, Dr. Tom Rosenthal has stepped up in her place. And he will be serving as the committee co-chair for the remainder of the project. So, I just wanted to publically thank Dr. Rosenthal for stepping up in Doris's absence and he will be helping Bruce facilitate today's call.

Dr. Tom Rosenthal: Thanks Ashlie.

Ashlie Wilbon: Thanks, Tom. Oh, and also wanted to thank Doris. I know she's not on the call but to thank her for serving as co-chair up to this point.

So, I did just want to recap briefly where we are at this point. Again, the goal this call is to discuss the scientific acceptability of the Ingenix measure. We did discuss the importance during the last call that we had on June 6, so I believe it's Cheri, are you on the call, from Ingenix?

Cheri Zielinski: Yes, I am.

Ashlie Wilbon: Okay, great. So if you could, well, I prompt you in just a second. We'll have you just do a brief introduction to get everyone back in the mindset of the measure. And then if you could also just - we did send out the information that you submitted in response to the question about the number of - the percent of claims that are ungrouped.

But if you could just kind of recap that response for everyone, that would be great. And I also just wanted to remind everyone that the HealthPartners measure that we discussed two calls ago I guess now. We did receive 16 total votes and by votes I mean actual recommendations for endorsement. We are trying to get a few more votes from the remainder of the people. It looks like some people went in and maybe got tired and didn't actually get to the end. So we're going to send out a reminder today. If we could get everyone else to submit their votes on that HealthPartners measure so we could get that one kind of wrapped up and closed out. That would be great.

So far, we are at 11 yes and - 11 for recommending for endorsement and 5 not recommending for endorsement. So and that's just a total of 16. So we still have about 7 people that could submit votes at this point.

And at that point, I think I'm going to ask if anyone has questions before Cheri jumps in and gives an introduction.

Okay, great, thanks. Cheri, could you just give a brief overview? Or Tom, whoever?

Cheri Zielinski: A measure overview of a - the responses to the questions, I'm sorry.

Ashlie Wilbon: Both, if you just, just to kind of reorient everyone to what the measure is and does. And then, just really briefly, and then also a recap of the response that you provided to that concern.

Dr. Tom Rosenthal: I would start with just the reason. Why is this here if you have ETG, something that's a background ((inaudible))?

Cheri Zielinski: Why is it here? Okay, so the measure is a population based measure. So it's not condition specific. It's providing metrics and measurement for population utilization metrics as a whole. Like, you know, resource utilization spend and ((inaudible)). And so it's - it is a accumulation of all of the conditions for a population, so it's member centric. And then the risk and resource utilization experience for those - for the population or the members as a whole.

Dr. Tom Rosenthal: I think a little different terminology from the last one but I understand, thank you.

Cheri Zielinski: Sure. And so the questions that were posed for follow up, there were three. And the first was, for those that do not group to an ETT are they completely excluded from the ERG? Or do they ((inaudible)) non-specific ETT or do they map into the ERG?

So before I go into that, I'll clarify and state that we incorporate another module here. And we - what we're trying to do here is which is ERG episode risk groups. What we - episode risk group is a population based risk assessment tool. And what it does is - it takes a look at each member and the conditions experienced in a one year period for each member.

And what we did was then we look at the risk score and utilization for all - the experience for one year's measures of episodes for that patient. And so when something does not it's called an ungroupable measure or ungroupable item. And so those are excluded - those are excluded from the - so if they don't group to an ETG they are not used in computing and ERG risk.

Jack Needleman: Can you give some concrete, this is Jack Needleman. Can you give some concrete examples of the kinds of charges that might not group into an ETG?

Cheri Zielinski: Sure, rule out lab and imaging services. Services with invalid procedure and diagnosis mapping wouldn't be groupable. Items like those wouldn't really have any useful information to ascertain risk anyway so they would be excluded from both the ETG and the ERG.

Jack Needleman: So some rule out diagnostic tests even they're part of a treatment says oh, we got you now. We now understand the diagnosis is not what we try to rule out might not get counted into the resources that we used in treating the patient or getting to that diagnosis. But it's a small percentage if I'm remembering the numbers correctly.

Cheri Zielinski: That is correct.

Jack Needleman: Both of those statements are correct.

Dan Dunn: Hey, Cheri.

Cheri Zielinski: Yes.

Dan Dunn: This is Dan Dunn. Actually I work with Cheri, part of the team, I'm supporting the measure. The mission to maybe get to Jack's question so think of it as to how does a service that doesn't group to an episode and therefore isn't used in calculating a member risk. And as Cheri mentioned that service won't, you know, find it's way into contributing to the risk score for that individual. However, that diagnostic service, you know, is part of the members overall cost. So it would be part of, you know, in the measure as a resource numerator, if you will. Even though it didn't contribute to the risk adjustment of that measure.

Jack Needleman: Okay, good, so we've got two paths. We've got the risk adjustments component. We've got the resource component and...

Dan Dunn: As we...

Jack Needleman: Resources are counted.

Dan Dunn: Yes, and it's because the numerator for this measure is overall costs. It's not, you know, costs as they relate to an ETG condition like some of our other measures.

Jack Needleman: Okay. Thank you.

Cheri Zielinski: Thanks Dan Dunn. And Dan Dunn will be helping to answer many more questions, so. The next question was what percentage of total costs do these missing records represent?

We ran a population of over 52 million claims in a data set for benchmarking and representing total costs of $6.9 million and of those, 2.6% of the claims and 2.67% of the total costs were deemed ungroupable. So a very small percentage.

And the third and final question was which type of episodes are they? We kind of already just talked about that with the rule out diagnosis and those kinds of things that are considered ungroupable.

So those were the questions we were posed for follow up to the measure. And that's the responses that we provided.

Ashlie Wilbon: Thank you, Cheri. I would like to ask, Carlos, are you there?

Carlos Alzola: Yes.

Ashlie Wilbon: Hi, so Carlos, for those of you that were not on the call last week, is the statistical consultant that we brought in by Inc US to help do some analysis of the scientific acceptability. Primarily the risk assessment methodology, the reliability and validity testing.

Can you just give a brief summary, Carlos, of what you found during your trials for the committee before they jump in?

Carlos Alzola: Sure.

Ashlie Wilbon: Okay.

Carlos Alzola: In the - in tests of reliability I have no issues. They look at the reliability of the data and they have elaborate system where they double code - although the steps that lead to the

production, to the generation of the data. And so they have to have a 99.9% match between the two approaches. And that's makes the data very reliable in my opinion.

And so that also affects the score because the score is part of this process. So I didn't have any issues in terms of the reliability.

Now in terms of the validity, I saw two things. The next thing is that they in the presentation, they say that they apply the methodology to self care data of different organizations but I didn't see in the submissions any detail results about that.

And, you know, so they tested face validity by intentionally modifying the data and converting resource pre and post modifications. But again, I didn't see any description of the results of that test.

And the also the recent attachment, the reliability validity testing but, again, there's a table there but I did not see any description of what those tables meant and how they related to reliability and validity. So I had those questions specifically about face validity.

My question refers specifically to the risk adjustment methodology. I see that they have like two different approach or very similar approach. Like the episode treatment groups, the ETGs, and the Gs. Now when they do the ETGs they assign a severity score. And it's a very complete methodology where they take into account not only the ETG but also culpability, state directions, between morbidities and the number of comorbidities that the patient has.

So that is, that seems a very complete model that effects all - that includes all the things that might effect the severity of the patient. And then they go to the ERGs that as far as I can tell is only a grouping of the codes - a grouping of the ETGs. And then assign again, different weights. So my question there would be what is the specific difference between the ERGs - the ETGs and

the ERGs? And why they thought that the ETGs were not sufficient for risk adjustment to the point that they needed to do the ERGs.

Dr. Tom Rosenthal: Can we take these in order. And Janis would you like to respond to the questions that Carlos raised about validity first?

Dan Dunn: Yes, this is Dan. Up to you whatever your preference is but happy to respond or follow up. Whatever works.

Dr. Tom Rosenthal: I'm suggesting that you address the validity issues he raised.

Dan Dunn: I think that one actually had some multiple pieces. Carlos, can you please maybe give them in order and we'll take them as they come?

Carlos Alzola: Sure. In the submission, you say that you applied the results of the methodology to the self-data of different organizations. But I didn't see the results of that process anywhere in the submission.

Dan Dunn: Okay and maybe just as a question does our combined submission include our follow up on scientific acceptability, Cheri?

Cheri Zielinski: I'm sorry we provided that with our responses, yes.

Dan Dunn: And there was a part of that, I think it's under SA1.3 related to testing results which did describe, you know, for example the individual R squareds of applying the ERG model to healthcare data of those customers. So is it something beyond? I'm not sure if you got that, Carlos. Are you looking for something beyond that?

Cheri Zielinski: I was going to ask if that was sent to the team, Ashlie, the responses that we provided?

Ashlie Wilbon: I'm not sure, obviously, we've been fielding a lot of emails. I'm pretty sure we forwarded everything. Can you just verbally if you guys have access to that to just tell the committee what that R squared was?

Dan Dunn: Sure, so this is - just to back up a little bit, this is taking the ERG risk score that comes from the model Cheri touched on a bit and we'll talk a little bit more about it to get to some Carlos later points.

So take healthcare data for those organizations, apply, you know, computer risk for each individual using ERGs. And the question is how well does that risk score correlate with the actual expenditures?

Carlos Alzola: Okay, that would be good that would be way to show valid.

Dan Dunn: And we did it using two different threshold truncations meaning we top code each individuals annual cost at 25,000 and then we did it at 100,000. Just because sometimes people use different thresholds, different applications.

And in the case of the - and actually just as another note - this was done to correlate with both total medical and pharmacy costs. And then total medical costs alone. So the question is again, take ERGs apply it, how well do you predict total medical and pharmacy costs? Second question, how well do you predict total medical costs alone?

And then for those two tasks, so medical and pharmacy costs with the $25,000 annual truncation threshold, ASCO is .57. And for 100,000 it was .55. These, again, are retrospective applications. So using a concurrence some people use that term.

So taking the ERG results for that year and see how well ((inaudible)). So, .57 and .55 and then the second model which is trying to explain medical costs variation was .53 and .52 for those two thresholds.

Dr. Tom Rosenthal: How many cases in those regressions?

Dan Dunn: Oh, god, millions.

Dr. Tom Rosenthal: Millions, okay.

Dan Dunn: Yes, I think it's seven million I think was this test. It was a lot and a mix of organizations in that pool of data.

Cheri Zielinski: Yes, seven million member sample, nine healthcare organizations.

Dan Dunn: Okay, yes, thanks Cheri.

Carlos Alzola: And do you take for - you said your ((inaudible)) rate is a percent of nine sensitive regions. Did you take one facility from each region?

Dan Dunn: Yeah and these are, essentially, think of these are health plans the healthcare organization. You know we also do these models and testing by organizations. We'll do it by geography. We can share those results with you, it all lines up pretty with the overall results.

Carlos Alzola: Okay.

Dan Dunn: It usually - if you calibrate off the one organization actually it gets a little better. So the joint one, if anything, is probably conservative estimate of what things might look like organization by organization.

Carlos Alzola: Okay.

Dan Dunn: We could follow up with that if that's helpful.

Carlos Alzola: Sure. My other question was I wanted to really understand what is the difference between the ETGs and the ERGs? Because I see that the ETGs includes a severity score and is correlated with resource use. But then you do a further a grouping - what it seems like is that is a further grouping of the ETGs into the ERGs. So I was curious about why you need to that? And what kind of coding uses the ETGs directly?

Dan Dunn: Good question.

Dr. Tom Rosenthal: Yes.

Dan Dunn: Maybe as a first comment is think of the purpose of ETGs as, you know, focus on episode based measurements. So for example an episode of diabetes or congestive heart failure, COPD and those severity models that are built separately for each one of those conditions, you know, gives you something to support risk adjustments, you know, for diabetes episodes specifically, for COPD episodes specifically, for CHF episodes specifically.

And the results of that are going to be going to be able to tag each episode for a member with not only the condition but a level of severity. So we know a member had a diabetes episode, a CHF episode, a COPD episode. And say the diabetes one is level whatever 3 - meaning a lot of things

going on that will affect thoughts. The COPD one is level 2, the CHF one is, you know, whatever level.

And that helps us to quote episode based measurements. And what E or Gs was designed to do is support member base, or population based, measurement. And basically what we're doing is taking advantage of what we learned from the episode grouping from individuals.

So we know their mix of conditions - I mean for other episodes. We also know the severity level of each one of those. And then to go from EREs to ERGs as you described, we're saying first when we see, you know, a diabetes level 3 episode. We're going to move it to this ERG which may be something like, you know, diabetes higher severity. And then maybe diabetes level 1 could be diabetes lower severity.

So when we're done with all that mapping because there's hundreds and hundreds of ETTs which can map down to a smaller number of ERGs which are, you know, maybe a little more of size and prevalence to support what we're trying to do.

After that step, we end up with a whole array of ERGs that are possible for an individual. So maybe, I think it's 160 or so of those to work with. So now we take an individuals mix of ETGs and their severity level. We can map that to the ERG markers designed for population based risk adjustment. And then, as you described, we're weighting each of those ERG markers to be used in our ERG score.

So maybe first comment in summary is ETGs designed for episode based measurement. ERGs for population based. And we're really taking advantage of the measurement to support the ERG focus.

Carlos Alzola: So, would it be fair to say that the weights of the ETGs are condition specific? But the in the ERGs you're just reweighting those - the groupings of people with different levels of severity or different conditions?

Dan Dunn: Yes, that's a exactly right. And think of the ETG models as helping us pocket the episodes by level. And then we're using that information on both what the episode is what the severity level is. That helps us do a better job with ERGs.

It's also ending up with that diabetes level 3, 4 mapping that to whatever it's diabetes E or G makers and then it's what you described. Where we're reweighting it for the ((inaudible)) purpose.

Carlos Alzola: Okay.

Dr. Tom Rosenthal: Any other questions about that relationship? I think that was something that other committee members have raised. The relationship between the ETGs and ERGs. And why go through the intermediate process of the ETGs?

Does anyone have any further questions about that?

Male: Yes, I've got a question about the role of the ERG weights which are still condition specific to an overall risk adjustment.

Dan Dunn: And concern?

Male: Well, it's not a concern yet. It's just a question. So, the ERGs are still diagnosis specific if I understand the documentation and your description of the role of. Is that correct?

Dan Dunn: Yes, there's some, you know, the ERGs a little less granular than ETGs.

Bruce Steinwald: Right. Okay, so my question is, okay, so now we're going to use the ERGs for risk adjustments. Trying to figure out what the overall cost of care for patient, to normalize that across different patients.

Now if I've got a complicated patient. Diabetic but also the diabetes has contributed to substantial vascular disease. Perhaps even, you know, and then unrelated to that in the same year that they develop melanoma. How does that get rolled up into a patient level risk adjuster? Are they additive? Are they averaged? Are they - is there some other mentioned weighting but is there some adjustments and that reflect the fact that we've got multiple conditions being treated potentially by the same physician.

How's that final risk adjuster developed from the ERGs?

Dan Dunn: Okay. Good question. Maybe two thoughts there. One is if you think back to how the ETG severity is assigned. That's assigned based on, you know, observing, for example, for episode of diabetes the fact that they also have congestive heart failure or CAD or some renal condition. So that will - in some ways that takes into account the disease interaction, diabetes with this other condition and you flip it back around.

So that those other conditions also can be affected severity wise by the presence of a comorbidity. So that candles the kind of the combinations of - and interactions between - conditions and ETGs actually gives us the basis for having that and taking advantage of it in the ERGs.

And each ERG weight is the incremental contribution of those markers. And then, as I think you were implying, you'll look at someone's ETGs and severity level mix. We'll map those to the ERG markers that they represent. There are also some hierarchies that are applied. So, for example,

you can't trigger a marker for both congestive heart failure and ischemic heart disease. The congestive heart failure will be the higher order condition.

But after that's supplied, an individual has a set of ERG markers. Each of those has a weight. So they're added up to create their overall scores.

Bruce Steinwald: All right, you just said they're added so if a given patient has a number of episodes in a year. Those episodes weights are added for that individual to come up with a final score?

Dan Dunn: No their episodes contribute to the ERG markers.

Bruce Steinwald: Right.

Dan Dunn: And so a member in ERG world will have - think of it as - say there's 150 ERGs. There's 150 toggles that you can flip on depending what episode you observed for them in that mapping. It doesn't know if you have multiple episodes for the same condition. It still only flips on one marker.

Bruce Steinwald: Okay.

Dan Dunn: As well as multiple related things. It'll still only flip on one marker, the one that's highest in the hierarchy. But after you've gone through that, every one of those markers you've, the toggles you've flipped on, has a weight attached to it and you sum up all those weights to get their score.

Bruce Steinwald: Okay. So the - sum up the weights to get score. So, again, I want to go back to my very sick patient here. Diabetes, vascular condition - what you're saying is those will in some way - because they're related to one and other - those will in some way be merged into an ETG or an ERG that represents vascular with diabetes complications or diabetes with vascular complications.

And that's a single score. The melanoma will be a separate score. It may be risk adjusted for the diabetes although it may not be given your algorithm but whatever the ERG for the melanoma is will be added to the ERG weight for the diabetes vascular condition to get the patients total ERG score for that year. Have I got that?

Dan Dunn: Maybe the only nuance is there'll be a separate market for diabetes and vascular.

Bruce Steinwald: Right.

Dan Dunn: And each of those markers will reflect the fact that both were observed.

Bruce Steinwald: Okay. And if they then go off to Africa and don't take care and turn malarial and they come back with malaria. We've got a third condition which is getting added to the ERGs - the total score from the three - from the now separate ERG associated with malaria will get added to the total score in terms of predicting their total costs.

Dan Dunn: Correct, yes.

Bruce Steinwald: And if somehow malaria was found to be useful. You know as a interaction with one of the other conditions in that malaria may actually change the marker that's related to those other conditions.

William Rich: Bruce, this is Bill Rich. Could I ask a question, please?

Bruce Steinwald: Absolutely.

William Rich: I'm going to address one of Carlos's notes on validity on 2Cs, the data representative of the target population. What just - do you feel that this can be applied to the Medicare population? And the second question is, was there any changes in the risk adjustment model for the underlying ETTs, where they were used as part of the building process?

Dan Dunn: That's, you mean, let me hit the first one and then I may need help on the second one. So your asking is a set of weights for some conditions that are - were designed specifically for the over - 65 and over - population?

And so there's a - for some of the ERGs - it'll look at age and not for all of them because all of them weren't found to be useful. But it will actually point to a different weight depending on age. Think of it as an elderly interaction with some of these markers.

So we, you know, a user is applied this model for, you know, elderly populations both commercial elderly and Medicare advantage and such. So the answer is yes, on that one.

Male: May I stop you there? Are you ((inaudible)) and what's the statistical basis for the Medicare population. Do you have, what's your sample that you're kind of weighting to in the Medicare population? I think that's...

Ashlie Wilbon: This is (Sally). I need to just interrupt really quickly because I know we talk a lot about applicability of the Medicare population for these measures. And as a reminder, because this measures only been tested in a commercial database per NQS policy it can only be endorsed for use in commercial populations.

So I know that it includes 65 and over members in the target population so I think that's a good point to pull up. But I just want to remind you all that this measure would only be endorsed for use in commercial populations.

Male: You answered my questions.

Bruce Steinwald: The second part was, was there any change in the underlying risk models in your ETGs over what you have published in your Web site a year or something.

Dan Dunn: No, the ETG models have not been - the risk models within - related to the ETGs have not updated or recalibrated in the last year.

Bruce Steinwald: Thank you very much.

Dr. Tom Rosenthal: Any other questions of this nature?

Carlos Alzola: This is Carlos.

Female: Well, this

Carlos Alzola: Yes?

Female: Go ahead.

Carlos Alzola: You have two risk adjustments a retrospective and a prospective one. The prospective includes the adjustment revision for age and gender. And the retrospective apparently does not. And the one you are applying is the retrospective?

Dan Dunn: Correct, for this measure, yes.

Is the question why the retrospective doesn't have age, gender ((inaudible))?

Carlos Alzola: Yes, that would be part of it, sure.

Dan Dunn: It ended up that they were pretty zero when we estimated the weights in the concurrent or retrospective model. Just weren't significant enough and just for, you know, simplicity they all got zero weighted in the end.

So the ETGs, you know, as mapped to ERGs ended up - there wasn't any residual that we could explain using age and gender.

Male: You have an explanation for that if you look at CMSs hierarchical groups, they had residuals or...

Dan Dunn: Actually in our prospective models the age/gender factors are in there. And they carry a decent amount of weight.

It's, you know, kind of related if you think, you know, in the year forward - even people without any clinical markers are likely going to have some medical expenditures. So most prospective models will find, you know, positive weight and include the age/gender marker the CMS model. Like the ERGs, like the, you know, the commercial ACC models.

Retrospective models though it ends up, you know, I think maybe - when you think, you know, the observed things that happen during that year pretty much explain, you know, the costs of that year. And it's - you actually aren't going to be able to observe any markers or costs for people who didn't have services.

Male: It's a typical finding with these. I, you know, think if you look at the other, you know, models of this type that people use. They've either left out age/gender markers or they have received a very low weight in their retrospective models.

Dr. Tom Rosenthal: Cheri or Ashlie, do you want us to now go through the criteria?

Ashlie Wilbon: Yes, that would be great.

Mary Kay O'Neill: This is Mary Kay; could I just ask one question because I've been thinking about this and maybe there's a simple explanation for it. But he said, ERGs are calculated for these individuals. Basically it sounds to me like you're looking at basically how many services were utilized and using that as a kind of marker for how the severity of the case of the individual. I'm trying to understand that we're not getting into some kind of circular logic when we're trying to measure appropriate utilization.

In other words, for example, people that have markers for surgical procedures for back pain. Would those guys look like they have a higher risk even if the back pain surgery was inappropriate? Simply because they received more intensive services?

Dan Dunn: Good question. Think of it as, so we're not using what was done. So whether a patient went in the hospital for back pain or has surgery for back pain or lots imaging. The only facts that we're picking up is the fact that back pain was present and it triggered an episode of back pain.

We're also looking - I'm not sure this makes sense for back pain - we're looking for other conditions. So diagnosis based conditions status in comorbidity. Factors that may make low back pain, you know, to have a higher expective resource cost.

But, so obviously you need to have - this is true of any of these risk adjustment methodologies - you need to have services to capture diagnosis's but other than that fact. And more services don't get to, you know, for the same diagnosis get to a different ERG.

But it's not in any way using utilization beyond that to compute risk.

Mary Kay O'Neill: So if you compared a population, you know, fairly large population in areas where there are known higher or lower utilization patterns for particular conditions. You wouldn't see the areas where there's higher utilization individuals looking like they're a higher risk group or members of a higher risk group?

Dan Dunn: If - not utilization, so if there's more utilization for a given condition...

Mary Kay O'Neill: Yes.

Dan Dunn: ...then you won't. You shouldn't observe any difference because it's based on their, you know, observed diagnosis rather what was done to diagnosis manage and treat those - those diagnoses.

Mary Kay O'Neill: Okay. Okay. Thank you.

Dr. Tom Rosenthal: Okay, you have up on the screen, those who can see it, the scientific acceptability criteria and sub-criteria.

Ashlie Wilbon: Yes, that's correct.

Dr. Tom Rosenthal: Okay.

Ashlie Wilbon: So if you want to - I think you probably discussed some of these. So, you know, as you go through you can kind of maybe recap or, you know, address, you know, just let everyone know that it's already been addressed. Or get a consensus on whether or not it's already been addressed.

Dr. Tom Rosenthal: So 2A1 is measure well defined and precisely specified. Any comments - now we're not, you know, we can go through and actually do our individual scoring of these separately. We don't have to do that now.

Ashlie Wilbon: Right.

Bruce Steinwald: But we're just going through them individually and giving people the opportunity to make a comment or ask a question.

Ashlie Wilbon: Absolutely, that's correct.

Bruce Steinwald: Okay. Anything on 2A1?

Male: I can't get into this.

Bruce Steinwald: Well, it's the, the measure is well-defined and precisely specified so that it can be implemented consistently within and across organizations and allow for (41). Electronic health record measures specifications are based on - well, it says the quality data set - but we know that it's something more in this context.

And then there are a whole bunch of sub-criteria.

Ashlie Wilbon: So Bruce, the items on the right side are just the specifications from...

Bruce Steinwald: Correct.

Ashlie Wilbon: ...the commission from the data applications.

Bruce Steinwald: Exactly.

Ashlie Wilbon: So, it's just - we're just determining whether or not those items that they submitted - the information they submitted for the items on the right, administrates 2A1, though.

Bruce Steinwald: Okay, comments or questions? While we move on.

Ashlie Wilbon: So Bruce, maybe, if we could just kind of get an idea of how the group feels about 2A1? Do they generally feel like that criteria was demonstrated and if not, maybe a rational or something. Just so we can - for our notes we need to have an idea of where the committee stands on that criteria.

Bruce Steinwald: Well, I think we'll accept silence as a mostly positive thing. So that if people do have concerns that they raise them.

Male: If I could, I'm not sure what the answer is, personally. But I'm not sure I would accept silence as agreement.

Bruce Steinwald: ((inaudible)).

Male: ((inaudible)) a mistake.

Bruce Steinwald: Yes, I didn't say agreement, but...

Male: Do we really have understanding about this in the context of the episode groupers?

Bruce Steinwald: Understanding - if you have a concern then please say more.

Male: No, I don't have a concern I'm just trying to be sure that in fact every-, that we in fact are somewhat on the same page about what we think - whether this is in fact well-defined and precisely measured so it can be implemented across within the cross organizations.

Bruce Steinwald: Does anyone want to speak in the affirmative around that point with regard to the measure?

Barbara Rudolph: This is Barbara Rudolph. I know a number of organizations that utilize this methodology and seem to do well with it. So, I, you know, it seems like it's well-specified, to me.

Phil? Yes, this is Phil Knight?. I still have a hard time reconciling Carlos observations under E and C inter-score with the answers we've heard. My silence is one of - I just can't be sure that all Carlos's terms were answered.

Bruce Steinwald: All right, this is Bruce. I'll withdraw the silence as golden remark but then, say, that if people want to make an affirmative remark certainly they should. If they want to express a concern, then they should do.

Joe Stephanski: This is Joe Stephanski. I am fine with 2A1. The risk adjustment methodology remains problematic for me on almost everything. I think we may need a little more detail at some point but right now I'd say, yes, we're precisely defined enough.

Male: And the detail is enough of, it's revealed enough that it could be reconstructed independent of Ingenix.

Bruce Steinwald: I think for the most part.

Male: Okay.

Bruce Steinwald: Okay so can we move on then? All right. 2A2 concerns reliability testing.

I don't know if I should read these or not. I know members have access to these from materials that were sent out in the past. And I'm reluctant to take the time to read everything. What's the view? Do you prefer that I did?

Ashlie Wilbon: I think for this one, this is Ashlie. I think for 2A2 which asks you to determine whether or not the reliability demonstrates the results are repeatedable.

Bruce Steinwald: Right.

Ashlie Wilbon: I think based on Carlos analysis, I don't know if anyone had any questions but it seems like their methodology for that in his opinion was acceptable. So maybe at least that can be a starting point for this one.

Joe Stephanski: This is Joe Stephanski. I think Carlos is right on this one.

Ashlie Wilbon: Yes.

Male: Good to agree with.

Bruce Steinwald: All right then. Then we move on to 2B1. The measure specification is consist with the evidence presented to support the focus of the measure under criterion 1B.

Measures specified to capture the most inclusive target population indicated by the evidence. And exclusions are supported by the evidence.

Ashlie Wilbon: Right, so this criteria's about whether or not, as set out, which is a population based measure to look at total cost if as constructed, the measure is consistent with that intent.

Bruce Steinwald: Comments or questions?

Barbara: This is Barbara I would say that it is consistent with the intent.

Joe Stephanski: This is Joe Stephanski. I agree with Barbara. I think we're going to have more trouble when we get down to the individual disease measures.

Bruce Steinwald: Okay. Can we move then to 2B2. Actually I'm having trouble reading on my own screen. So and Ashlie, why don't you take over on these please.

Ashlie Wilbon: Okay, so 2B2 is asking whether or not the submission demonstrated that validity testing for the data elements are correct and the measure score correctly reflects the cost of care provided. Adequately distinguishing higher or lower costs or ((inaudible)).

Male: And as a reminder based on NQF recent guidance what is minimal acceptable for this sub-question, states validity and a measured developer can demonstrate validity either by showing that the data elements are valid that are used calculate the measure or the measure score or both.

So there are many levels in which they can meet validity testing.

Bruce Steinwald: Okay.

Male: Any comments on this?

Bruce Steinwald: Comments or questions?

Male: The material that was sent out suggested though that this had been - that this had not been - I mean validation can be against other data sources or I guess against this great phase test. This has been tested against the straight face test but it has not been tested against other data bases because it's not feasible to do that, I guess.

Male: There's not a lot of databases to compare it against.

Male: Right so we're only left with the straight face test of does this look like it would be accurate and valid in what it purports to be measuring?

Bruce Steinwald: Comments, any further comments or questions. Okay, let's move on.

Ashlie Wilbon: So Bruce, again, or Thomas, if you guys could just - we, kind of give us a summary. Maybe even kind of going back to Carlos's questions and how Dan and Cheri responded to his concerns about the validity. The measure.

Dr. Tom Rosenthal: Bruce, I'll let you do that one.

Ashlie Wilbon: Perhaps the follow up information I don't know if they were able to provide the R squared. I don't know if we ended up getting a number on that or not. But if there's additional information that you guys would like - you guys being the steering committee - in order to feel like you're comfortable with rating this sub-criteria, let us know.

Male: I've got a question for the Ingenix people on this issue. Since this about we basically getting the data and it talks about commercial. You've got both fee for service and HMO database. Are you -

how much - fee for service guidance, we're always dealing with billed elements here. Fee for service, the incentive is there to bill, are you confident you're getting full billing information from your groups or physicians that are being paid on some basis other than fee for service.

Dan Dunn: Yes, great question. I think that kind of gets to ((inaudible)) made a point. Like all of our testing and development data, we don't include any data sources where, you know, capitation or carve outs or other things that would result in the missing information that could cause issues or order.

But the other side of it is, if you're a user applying this measure, you know, our guideline is that if you do not have, you know, full information for a population or any part of it. Either exclude that part of the population where is missing information or if, you know, if you have issues across the whole dataset then you really can't apply this measure.

Well, the exception is where pharmacy data is not available due to benefit coverage or something like that - this model actually adjusts for the presence or absence of pharmacy data

And just one last though, there are users of this measure who when they have, you know, capitation for example, arrangements. They do collect the encounters and are able to price them. But that's been on the user's side for them both to - will give guidelines but they then decide whether they think they have complete data or not. And where they don't have complete data, whether they've captured it sufficiently with encounters and, you know, costed them out appropriately.

I'm not sure it answered the question. Maybe it's the bottom line. This isn't going to work well like a lot of these measures unless you are confident you have, you know, complete and consistent information to support them.

Dr. Tom Rosenthal: Well the two big carve outs in a lot of commercial coverage are as you noted pharamacy benefits and mental health, behavioral health, benefits. And in the validation sample how did you handle getting information on both of those, what percentage of the claims that have come to your - even your fee for service - commercial clients included data on those.

Dan Dunn: Yes, and pharmacy data in both, not all validation, but in all development. We actually work with a population that has pharmacy data then we pull it out to develop kind of the medical version of it. So it allows us to develop the model to cover both those scenarios.

The case of mental health benefits that is something where we'll exclude, you know, data from our development side, our testing side, if it has a carve out for mental health benefits. We just won't use the data. It comes back to the customer. If they're maybe outside the facility - there's no adjustments directly in the measure for absence of mental health service claims? And so it's up to the user of the measure to ensure that they have that information available.

Dr. Tom Rosenthal: So in terms of the application of the measure what's been your experience working with clients that have carve outs in both of those areas in terms of generating total resource use for beneficiary measures that accurately reflect the full resource - full billed resource use.

Dan Dunn: Well the pharmacy one hasn't been an issue because, you know, that's again up to the user whether they want to fully include populations where you have pharmacy as well as medical claims. Because the methodology will adjust, so you actually can have a mixed population, benefits and pharmacy. And you pretty much measure what you have and it will do it accurately. And that's a common approach, a lot of it. And, you know, maybe they'll measure medical cost separately from medical cost pharmacy to allow to take advantage of both but be able to isolate the medical piece if they want to.

On the mental health carve out, you know, there are likely users who will apply the methodology even in the absence of mental health. And so that effects two things. One is, you know, it's not going to fully - because you're not going to get the mental health claims to trigger the ERG risk score. I mean it's a little bit understated.

But the other side is when people use these measures, you know, if you go to the details of how the measures are described. You know, usually they're using some kind of observed and expected general approach. And essentially when they're doing that they're creating that expector which is in a sense an internal benchmark based on their own experience.

So if they've consistently across the population have that mental health carve out. A lot of them have been comfortable with the fact that the, those internal benchmarks where left out and the measure will still work. But that's up, you know, up to the user to - my sense is it works pretty well even with the absence of mental health benefits given that going to backend adjustment.

But the measure itself doesn't judge for that specifically.

Bruce Steinwald: Okay.

Ashlie Wilbon: This is (Sally). Just a question for the steering committee as we continue to move forward and looking at other measures that is this not an issue that touches on all measures when we're talking about carve outs in the commercial population.

We're going to jot that down as a note to circle back to.

Bruce Steinwald: It potentially could pertain to all non-condition specific measures, yes.

Ashlie Wilbon: I'm just thinking about the application of this measure from our perspective being something other than a sort of internal checklist within a given organization. And I could understand within a organization, you know, they're doing that comparison with themselves over time on the same data inputs or data exclusions, whatever but if we're starting to compare across a community or whatever using these measures. Then all of these things really need to be carefully teased out. And I would think we would have a problem if we started missing diagnostic data, particularly on mental health if we're trying to really risk to get them.

I guess that's how we're planning on using these measures which I assume is in somewhat of a new way as to the original purposes they were developed for earlier.

Female: This is ((inaudible)) I think these are very interesting questions. And what it strikes me are they - are we - is the generalized ability of the measure knowing that data out there are missing? And are missing in different amounts or different types across communities? I think you're right, we need to tease this up and think this through. But also challenging ourselves to think about where the implications for the measure could be endorsed if they are not considering these things? Or, you know, how would we recommend that developers account for these types of differences?

Ashlie Wilbon: And in the same thing using this measure in a community setting. We wouldn't really be able to compare resource between a HMO and a classic fee for service organization based on what we just heard either. Right? Those wouldn't be comparable entities.

Bruce Steinwald: I'm sorry why do you say that?

Ashlie Wilbon: Well, I thought that there was - if you were paid on a capitation basis that there was not - that we couldn't use that data. I mean so, like, in the Seattle example we really couldn't compare the effectiveness or the efficiency of group health versus the rest of the community.

Bruce Steinwald: Well, if you're able to measure the same resources on a per population basis and have to exclude, let's say behavioral health resources. You would still have comparability between two populations.

Ashlie Wilbon: Sure, yes, but this back to an earlier point on the use data that, you said, I thought that (Bostram) and Ingenix said that we couldn't use capitation because those were often not complete data sets.

Dan Dunn: This is Dan maybe to help. I think the general principle is if you're missing data whether it's due to a conformation carve out where you don't get the encounters. Or even the behavioral health issue, the capitation to be honest is a more significant issue. But you have the HMO population that has all the primary care and specialty services capped out to some physician arrangement then you really can't support this measure.

So I wouldn't even go as far as saying you can't compare them. I would say that you don't meet the threshold of data consistency and completeness to support the measure.

Bruce Steinwald: But this would be true for any population measure that we're likely to grapple with.

Dan Dunn: Yes. Any condition one, as well, even the mental health thing, to be honest, except the condition based ones. Because most condition base will use, you know, depression and other mental or behavioral health disorders as something that affects the severity of the condition.

Bruce Steinwald: Great.

Female: But you, could perhaps pick up the mental health diagnosis from the primary care encounter or something like that. Right? So, I mean they don't have to get mental services within the data that you're looking at to get that risk assigned ((inaudible)).

Dan Dunn: Yes, you're right. So you get some of them likely, you miss some as well.

Bruce Steinwald: It's all due to some real peculiar situations in the data and I've got to admit at this point I'm probably on my level of knowing how mental health is handled in these carve outs. But an awful lot of depression care, for example, is delivered by primary care docs with an awful lot of prescribing of psycho, you know, drugs for depression.

I don't know whether those visits wind up in the mental health carve out or whether it's only if you're going to a specialized mental health provider that you wind up with a carve out. So now we've got some issues of is the primary care doc doing the treatments for depression. Have they referred to somebody else? Is the somebody else in the carve out?

Now we've got issues are the - not only the full resources for behavioral health in the data base for examining the total cost of care but we also have the issue that we've got a primary care doc doing some mental health services. So those costs are in there. And so depending on the physician preference for referral versus treatment for certain kinds of mental health. We may or may not have those in numbers in and we've got inconsistent data.

Ashlie Wilbon: And this is (Sally) just to remind all of you because this conversation I think was touched upon when we though about what we wanted in terms of specification submitted to NQS. And for the resources he measured, we did request that the measure developers submit their data protocol components, which includes instructions to the users on what type of data they require. Whether other kind of steps to deal with missing data and Ingenix, also, all the measured developers filled out that piece.

So exactly because of these concerns that all of you had, unlike for some of the other measures. We did request this information as part of the submission process. So just as a reminder to the group.

Dan Dunn: And this is Dan. I think when you asked us for guidelines around, you know, data consistency and completeness our response was other pharmacy which we had specifically adjust for in the measure. You know, the assumption is that everything including behavioral health claims are available and are being used.

Dr. Tom Rosenthal: Okay.

Bruce Steinwald: Does that mean we should move to 2B4?

Ashlie Wilbon: Yes, I guess.

Bruce Steinwald: I mean it sounds like we just discussed 2B3 to a large extent.

Ashlie Wilbon: Yes, yes, you're right. So, yes, 2B4, I'm scrolling down for those of you watching the webinar. It takes us a couple of seconds sometimes to catch up.

Female: 2B4 is about the risk adjustment methods.

Ashlie Wilbon: So, in case you're waiting for me to read this. It's asking for whether or not there was an evidence based risk adjustment strategy. If specified and based on clinical factors. Or that the risk or that if they didn't have risk that really doesn't really apply, so.

Bruce Steinwald: Right.

Ashlie Wilbon: Whether or not if they had a rationale for no risk adjustment. But they did have risk adjustments.

Bruce Steinwald: Well, we certainly did discuss this. Would anyone like to raise any lingering concerns? Or make a positive statement?

Male: Sounds like this one's been beat to death.

Bruce Steinwald: Yes, I think we talked about it enough.

Male: So.

Ashlie Wilbon: So just a quick question from staff again, this is Ashlie. If we could just get a general comments on whether or not you are comfortable with the risk adjustment method they used. I know there was some discussion from Carlos and from Dan and his response about the ERG versus the ETG risk adjustment methodology and whether or not the committee feels that that was adequate?

Bruce Steinwald: Well, Ashlie since we all have to score the all the criteria why do you need a sense of the committee since you were trying to get all the results within real time anyway.

Ashlie Wilbon: Right, so even though we get your ratings, we do need a rational. And we summarize the meetings and the evaluation of the criteria with your discussion and rational for each of the sub-criteria. So, right, and we also want to make sure we don't need any follow up from the developers.

So make sure that you guys have what you need to submit the ratings on line.

Bruce Steinwald: Well, let's address that then. Does anyone feel like we need more information from Ingenix? On risk adjustment.

Male: Nope.

Bruce Steinwald: I'm reasonably satisfied.

Dr. Tom Rosenthal: And can we summarize what the general sense of the concern was? Would the general concerns had been addressed adequately?

Bruce Steinwald: Well, what do you think?

Dr. Tom Rosenthal: I don't know. I missed that part of the conversation.

Bruce Steinwald: Oh.

Dr. Tom Rosenthal: So, I'm asking. I'm asking and I'm trying to re-channel Ashlie's request of us.

Bruce Steinwald: Well, it began with Carlos asking for an explanation of the risk adjustment and then the Ingenix staff basically providing an explanation and some steering committee members asking questions. To my mind it was dealt with adequately. But, please others if you have lingering concerns. I remember someone said that this issue might more of an issue with the conditions specific. But, please, Ashlie and Sally, if you don't have enough information don't let us go on until you do.

Ashlie Wilbon: I can summarize what we think we heard from the group.

Dr. Tom Rosenthal: Try that, try that and let's see if that works.

Ashlie Wilbon: So, we had some questions about exactly what the relationship is from ETG to ERG and how it plays itself out with some very concrete examples. And my sense was that it was adequately described by Ingenix back to the group. I didn't hear much follow up questions after that conversation.

So my sense, what we're hearing is that the risk adjustment appears to be adequately described. Though what we're not sure about is whether or not any of that follow up needs to be in writing or if the meeting summary is enough. So we've kind of heard after going back and forth that the risk adjustment seemed okay.

I don't want to say moderate or high or whatever, that's up to you guys but we didn't hear anything that signals us that it was going to not pass at all.

Dr. Tom Rosenthal: So that's the summary that you would in effect try to take away from the meeting. And does that jive with people's general sense of what was discussed?

Male: Yes.

Dr. Tom Rosenthal: Okay, there we go. Okay. Let's move on.

Ashlie Wilbon: All right, so, 2B5 is the next ((inaudible)) that asks whether or not the data analysis demonstrates that the methods for scoring and analysis of the measure allow for identification of statically significant practically or clinically meaningful differences in performance.

Female: I guess I'm still trying to work though - the general application of these measures beyond the walls of extend facility with the variability of the inclusion of pharmacy and behavioral health data. I mean I understand statistically why in one situation you include it and in another you wouldn't.

But then I'm just trying to figure out if you're going - if you're applying this more broadly than within the walls of a given institution. How, you know, how that might affect the practical significance of the measure?

Ashlie Wilbon: Dan, do you want to maybe respond to that? Dan or Cheri?

Dan Dunn: Is the question that you're comparing different organizations within a community where there's differences in data available to support the measurement? Is that the challenge?

Female: Yes, yes, so I mean so if it's an ETG with and without behavioral health data because of the, you know, the nature of the - you know the nature of the data availability. How would that work? I guess, this is pretty complex. I guess in that community there'd be different payer mixes and the availability of that data may vary more payer by payer than it would institution by institution.

Dan Dunn: Yes, I think I'd make a general statement. Like, maybe I shouldn't but I will. Anyways but, any of these resource use measures I don't think that's - I think there's a way to adjust for with or without pharmacy because it's such common sort of data difference in most measures including ours will actually calibrate differently depending with or without.

But if you get into things like capitation, information missing due to that or behavioral or the lab claims aren't showing up because that's carved out to Lab Corp, somebody. I don't think that's an endorsed use of the measure.

I probably shouldn't use that term. But you're right you can't make valid comparisons across organizations with different sort of data completeness and consistency.

Male: I would also like to point out as a economist, carve outs create very different incentives for practitioners in terms of how they practice. We have some research not enough on how strongly the incentives are matched. But if by referring somebody out from my capitated plan to my physician group capitation to behavioral health, I shift the cost to a different capitated element.

Or I make use of drugs rather than talk therapy because drugs are covered by somebody else's plan. I've got real different incentives where there are carve outs. So understanding total resource use and how resource is effected by the incentives is something that's very hard to understand without having the total costs across all the different carve outs.

And that behavioral response is going to be very important as we think about the impact of ACOs and other types of organizations down the road as we try to change reimbursement patterns.

Male: You know, that's a great point because it's not only the missing data it's what you see in the data that remains if it's an issue as well.

Male: To read 2B5 closely it says that the data analysis demonstrates that the methods for scoring an analysis allow for identification of statically significant and clinically meaningful differences in performance.

Could the developers give a couple of specific examples of how they believe the measure meets that criteria?

Dan Dunn: Sure. I'll give it a shot. So maybe design or by observations of use in practice. Because by design, you know, the approach to measuring risk, you know, is designed to in many ways equalize the comparisons across, you know, patients and across groups of patients that may be affiliated with the physician.

So if the risk model works well, that'll, you know, remove differences related to difference in underlying, underlying morbidity of the population. And then the resource use measures are fairly straight forward in terms, you know, of measuring what you have. So I guess, by...

Bruce Steinwald: Yes, I guess by design, my interpretation of 2B5 and maybe I'm misinterpreting it. Is that this is the practical side of the thing. Does it in fact, given attributions and groups, etcetera. Does it - does it in fact distinguish statistically diff-, significant and clinically meaningful differences. In reality.

Dan Dunn: Yes, no, if you look at users of this methodology and say they're comparing primary care physicians or their comparing delivery systems. My experience is when they've done that and they've found a physician or a delivery system to be different. Specifically both magnitude wise and statically that when they look behind the results, you know, they can usually identify reasons why that's the case when they talk with those organizations.

You know, it's usually around why the differences are seen and what can be done to improve. If that helps, but that's probably the - in terms of - you do see distributions of performance here. You do see some entity being measured being statistically different than a benchmark or the average for the whole group. And when you look at the details behind the results, you know, it does, you know, in general seems to be basically of why that group or physician was different.

Bruce Steinwald: Okay, any other questions or comments on this point?

Ashlie Wilbon: And this is (Sally) just for those on the webinar we did pull up the example report that Ingenix supplied for this measure. And if you are not able to get on the webinar, it's the S12 underscore sample underscore in your PDF packet. So they did provide examples of how it might be practically in real life reported.

Ashlie Wilbon: For 2B6 I think for this measure it's using all administrative data so we can, generally been counting that one in A as not applicable. I'm sorry, I guess I should ask if there are any more comments or questions on 2B5 before move on.

Bruce Steinwald: I got the impression we were good on that.

Ashlie Wilbon: Okay, okay, so 2B6 will be not applicable. And the final criteria for scientific acceptability addresses disparities, whether or not their disparities in care had been identified in the specifications and whether or not they have identified disparities through stratification of the measured results.

Bruce Steinwald: Any comments on this?

Female: What administrative data set in it's current condition - other aspects would not be identifiable. Is that correct?

Male: So they don't have data on ((inaudible)) I don't believe.

Dan Dunn: Right so we're and if we had it - this one usually gets handled by the user, stratifying their population in terms of measurement. We don't include any risk or other adjustments based on ((inaudible)) and gender of the ((inaudible)) because if someone mentioned we don't have the ability to pull this data on a continual basis.

Bruce Steinwald: So the answer to this no...

Male: That's right.

Bruce Steinwald: ...and the question is does it matter to any significant degree?

Dr. Tom Rosenthal: Right, does it matter as a measure, I mean,...

Bruce Steinwald: Right, these would be the measure, that's what I meant.

Dr. Tom Rosenthal: ...the fact that, yes, yes. The fact that the user can do it if the user wishes to is not really germane, I guess to the measure itself.

Bruce Steinwald: I'm not even clear how the user - how the individual user could use it if the race/ethnicity etcetera is not captured.

Dr. Tom Rosenthal: That's exactly right, right.

Bruce Steinwald: And the answer's still no. And the question does that possibility that these factors aren't considered. Does that impact positively or negatively on accepting this as a measure? I would suggest not, but I'm just. That seems to be the question that 2C poses.

Male: Specific question for Ingenix. Doesn't ACA mandate the collection even in the commercial world of race and ethnicity?

Dan Dunn: Good, good question. We haven't...

Dr. Tom Rosenthal: Well, it does actually.

Male: Yes, okay, but if the question do we end up - does it end up collected and whatever, integrated into the type of data that supports this. We haven't seen it come through. So maybe in the future this could be handled but not right now.

Ashlie Wilbon: And from a public health perspective if you could collect it would explain some risk adjustment I would assume but I don't think anybody's really collecting it now because the systems not there. Right? I mean it's a hard piece of data to collect.

Male: ((inaudible)) care does.

Ashlie Wilbon: And this is (Sally). Just to add a little bit of perspective on NQF purpose of this sub-criteria. It is that we do not want to adjust away disparities that should be exposed. So 2C is really about if there are disparities are they being just pulled into the risk adjustment so that they're less likely to be actionable.

So in fact for the quality measures where this really came from it was when there are, for example, racial disparities in cardiac care. That those are not being risk adjusted away but rather there's a recommendation that the users stratify by these so that they can adequately and appropriately take action to reduce those disparities.

Bruce Steinwald: Well, I think in the - I get that. And in the cost realm of the thing, at least as I see it. The question would be slightly answered - or asked or answered differently which is to the degree that somebody in a low socio-economic class/status either uses more resources or more resources as a result of being in that socio-economic class. That would be a relevant factor.

Female: Yes.

Bruce Steinwald: Potentially and if it's not captured, you'd have no way of assessing it. And in most of these administrative data sets that information isn't captured.

Male: And where it is captured it's usually inaccurate.

Male: Right.

Dr. Tom Rosenthal: I personally don't think it matters with regard to the Ingenix thing because these are basically commercial populations where there's a somewhat homogenous, although I may be - closely overstating it. But in the range of realm of incomes and socio-economic statuses are likely not to be playing a factor.

Male: But if ((inaudible))...

Jack Needleman: Tom, on that point I'd just like to point out that at UCLA where we both worked. At the hospital that you worked in, you've got janitors and radiologists in the same - potentially in the same commercial group.

Male: Right.

Dr. Tom Rosenthal: I'm happy to be stand - stood corrected. I'm...

Barbara Rudolph: I think there's much ((inaudible)) what I find is people in public health think that there's much more homogeneity in the commercially insured population than there in fact is and there's actually quite a range.

Dr. Tom Rosenthal: Well, that was what Jack just said, in correcting me. But, again, do we think that that's relevant in terms of either accepting this Ingenix measure or not because it appears to not be captured.

Jack Needleman: Well, I think it's important that we note it.

Dr. Tom Rosenthal: Right.

Ashlie Wilbon: I think we have note that it's not - that we can not capture it currently, systematically I don't believe.

Bruce Steinwald: And I will point out there are administrative databases where you can pick up striking differences in utilization on race and ethnicity. So someone said there weren't but there certainly are.

Dr. Tom Rosenthal: And you can mistake someone being efficient for actually someone who has a disproportionately minority or non-minority population in their patient population.

Male: That is so.

Bruce Steinwald: I think, Jim, one of the things we're going to have to grapple with as a group is most - I'm overstating these things - a lot of the measures that we're likely to see are not accurately capture any of that. And are we going to make it such that that would render a particular measure non-supportable on the basis of that alone.

Dr. Tom Rosenthal: Still I think it should in the outfit, all of the commercial we're going to see, this is going to be a defect in them as of today. Hope that they'd be captured in the future as ACA, the law...

Male: Okay.

Ashlie Wilbon: Yes, I think I think that's going to evolve. We should have it as a goal to have it. But I don't - we shouldn't use it to stop something today.

Bruce Steinwald: I concur.

Male: All right.

Dr. Tom Rosenthal: So (Sally), I think we got the answers on that one.

Ashlie Wilbon: Okay, okay, thanks Tom. And just as a staff suggestion based on your discussion now, we'll be noting that discussion in our notes. And based on that we probably recommend that for measures that you feel is applicable that we would rate this not applicable.

Male: Right.

Ashlie Wilbon: In the criteria rating.

Male: In the criteria rating.

Ashlie Wilbon: But we will in the report even include this conversation about the adequacy or inadequacy of the commercial data. And that it is desirable avenue to go down but given where we are now, the measures aren't expected to address it. If that sounds right.

Dr. Tom Rosenthal: I would support because say it's not applicable is not appropriate. There are - when these things are used on an individual level, not a population base, they can have - someone mentioned - they can have huge disparities or implications for resource use.

Bruce Steinwald: Well, and they can have the same implications at a population level.

Dr. Tom Rosenthal: Exactly. Well, that's correct.

Bruce Steinwald: Right, yes, so we don't want to say it's not applicable. We want to say whether or not this is determining of whether we accept this particular measure or any other measure.

Ashlie Wilbon: Okay, that's it.

Dr. Tom Rosenthal: And, again, even within the context of the measure which isn't to deal with the whole issue of health disparities per se but simply to compare physicians - or not simply - but one of the major goals physician within a single group.

If the janitors and radiologists are going to different physicians, then this issue socio-economic adjustment becomes an issue if we think that those. The underlying socio-economic status of the patients affect what the physicians can do or the resources that they can put in place. So it's in that context we need to think about - that's the first context we need to think about the absence of SES data in the stratification or the ability to risk adjust.

Bruce Steinwald: A very practical level with as we're getting - folks are being, forming larger and larger groups. It's different offices will have dramatically different socio-economic makeup and the physicians that are in those offices going to have markedly different profiles with that group.

Ashlie Wilbon: I think we've got what we need.

Ashlie Wilbon: Yes, so.

Bruce Steinwald: Okay.

Dan Dunn: This is Dan. Could I (get a quick one), has this discussion come up on the quality side? Quality mea-, I would think it would have applicability there as well.

Ashlie Wilbon: Yes. Yes this discussion comes up in quality measures quite often, Dan. And I that the path that NQF has been on is kind of signaling for measures to start including how you might stratify the measure by these disparities even if the data aren't supporting it.

It's doesn't - I think the steering committee is right on target of what we've heard in the past. It's not going to enough to bring down a measure but it's important to send out a signal and let people know that this is something of interest.

Dan Dunn: Okay, thank you.

Bruce Steinwald: All right, (Sally) what's next.

Ashlie Wilbon: So at this point, it sounds like you guys have gotten through all the sub-criteria. For scientific acceptability and we have what we need I think at this point in terms of having a rational for where you stand at this point prior to completing your final rating.

You have a choice at this point. And I'll allow the co-chairs to pose this to the group. We can try to get through usability on the call, we have another 30 minutes. Or we can stop now, 30 minutes early and carry over the usability and feasibility discussion of this measure to the in person meeting.

So, keeping in mind that, you know, if we add it to the agenda for the in person meeting we may cutting time for discussion on other measures but just wanted to put that out there as an option, so.

Dr. Tom Rosenthal: Well, we have a lot to do at the in person meeting. Do you think, well I'll poll the group. Do you think that we should address these issues and at least learn whether there is a need for lengthy discussion or not?

Bruce Steinwald: I would say so but I must tell you that I was planning on using the webinar because I left the materials - I'll look at them this weekend and the webinar I can not get on.

Dr. Tom Rosenthal: We have technical problem of being able to actually go through the usability/feasibility discussion.

Jack Needleman: I've been in the Commonwealth of Virginia today, Bruce.

Bruce Steinwald: You know that happen, I tried again and I was able to get on, so.

Male: I tried about four or five times.

Jack Needleman: I was kicked out once and got back on.

Dan Dunn: Ashlie before, can I just ask, does the issue about attribution, does that come up in usability and feasibility? Is that were that gets addressed.

Ashlie Wilbon: A little bit in usability perhaps and how the scores are reported or how useful the information is.

Dan Dunn: Okay.

Ashlie Wilbon: But I think a lot of that would have been addressed in the specifications and how they - in the reporting module of the specifications and how.

Female: ...the claims.

Ashlie Wilbon: And then validity. Whether or not how they recommend attribute is a valid approach given what the intent ((inaudible))

Dan Dunn: Okay.

Ashlie Wilbon: Right?

Ashlie Wilbon: So usability we could just ((inaudible)) to help you make a decision for those of you who have not been able to get on the webinar. We apologize. We also understand there's issues with the phone with people getting kicked off. We will be having - we will be following with our meetings people to report these issues.

But usability looks at whether or not the performance results are being used currently. It looks at whether or not the results themselves are meaningful, understandable, and useful for the intended audience. So it does take into account who the intended audience are.

That the data and the results detail are maintained such that if they need to be updated they can be and whether or not they're transparent enough to facilitate understanding particularly by the ((inaudible)) who are being measured.

And then, and that's it. Those are the three sub-criteria that apply to usability for the ((inaudible)) across the measures. So to maybe help you decide if you want to use the next 20 minutes to try to go through those.

Bruce Steinwald: This is Bruce since we have the Ingenix people on the line. I'd like to hear a little bit about how the measure actually is used and how one obtains the rights to use it. Could you say to just give us a little overview of that?

Dan Dunn: Sure on the first one, most - give you two of the applications that you see the most.

One is to measure primary physicians around, you know, the management of a population. I don't think there are many cases I can think of where I've seen a specialist physician or a group measured using this approach. So if it's a medical home or some type of model where there can be some, you know, was that the question? Sorry.

Bruce Steinwald: Yes, how is it being used? Yes, that's the question, how is it being used?

Dan Dunn: Okay, was I addressing it or missing the...

Bruce Steinwald: No, I think you were headed in the right direction.

Dan Dunn: Okay.

Bruce Steinwald: But who uses it? It is it health plans that use it, is it hospitals that use, is it ACOs that use it? Who's currently using it to hold primary physicians accountable either individually or collectively?

Dan Dunn: Health plans, ACO Lite organization - I'm sure if there's any ACOs out there yet - but, you know, delivery systems who have some interest in this ((inaudible)) area, cost of care resource use. You know, anyone who, you know, has an interest in identifying opportunities or using it to measure entities, you know, for information sharing or even, you know, incentive based payments.

Bruce Steinwald: Can you give us examples of which those organizations are or is that held close?

Dan Dunn: Well, on the health plan side, you know, pretty much a, you know, good cross section of all the health plans in the US are using ERGs. Not every single one uses, you know, some of them will use other similar, you know, risk adjustment methodologies than ERGs underscore population measurements.

I could name almost any health plan at least dabbling in this for either information sharing or even those incentive based payments on the physicians side. I think I can name organizations, like (Suder), (Cleveland Clinic), you know, some of the - I'm in Boston; some of the delivery systems in Boston are using it as well.

Bruce Steinwald: What about employers any?

Dan Dunn: Haven't usually see - you will see employers do things like measure their populations using this methodology against benchmark populations or if they split their population in some way by geography, just suggested comparisons. But not, I haven't seen any of them doing specifically provider based measurement using this methodology.

Bruce Steinwald: But they could compare if they're multisite employers, they could compare performance at different sites using this measure, couldn't they?

Dan Dunn: Yes, exactly. Yes, that is, that point, right. They'll cut their population up and, you know, they have comparable risk measurements, they have comparable resource use measurements. They can, you know, apply and compare one against the other.

Bruce Steinwald: And the other question was how does one obtain the use of the data? Do you license it or do you do all of the computations yourself under a contract with your clients? How does that work?

Dan Dunn: It's different models. There are models where maybe as a two set thing here. You know, this is a similar case with some of the episode based methodologies including ours.

Most users, most all of them, will license a piece of software from Ingenix. That helps implement the methodology because some of this is a good amount of work if you wanted to, you know, kind of replicate from the details.

So that part is typically licensed no matter what the ((inaudible)) so that will give you the risk piece of the measure. How people then take those results and create kind of the final resource use measure. There are some cases where we'll work with organizations to apply and they'll send us their data, we'll apply it.

There are other cases where the organization is license, you know, ERGs and they're doing their work themselves.

Bruce Steinwald: Okay, can I ask you a question relationship to the usability in the following way. Would it be an accurate comparison to say Blue Cross of Ohio that if they have a spectrum or primary care physicians that have a total cost of care of some number. Is that number accurately comparable to say a primary care group at (Suder) in California?

Dan Dunn: From the - assuming they both have complete and consistent data...

Bruce Steinwald: Yes, yes.

Dan Dunn: ...the risk core should be comparable, you know, if they standard price, if they have a similar benefit structure at (Suder). If they standard price their resources or their services, you should be able to compare. You know, the same way if you look at what MCQA is trying to do with the RRU

measures. You know, they've been designed to support that type of comparison if you can assume you've got consistently priced whatever resources to do the comparison.

Dr. Tom Rosenthal: So is there an example of a current user who's done something between regions like that or is it pretty much all within a single region. The examples that you gave of current use.

Dan Dunn: Around physician measurements, most organizations will, you know, assume that within a community and maybe the broadest would be within a state. When they create sort of a pool of focus to do comparisons. I'm sure some of the larger, larger national health plans are probably doing comparisons across, you know, one state plan versus another using this type of methodology.

I don't think a lot of them are - even though they may apply for physician measurement, apply the same methodology across. They usually only will pool the comparison locally.

Bruce Steinwald: And will you articulate again, if you don't mind, how in fact you create the standardized pricing? Because there's not only risk adjustment that's required for this but there's clearly standardized pricing that required for this for the comparisons to be meaningful across regions.

Dan Dunn: Yes and right now I'll be specific we did not submit a standard pricing methodology as a part of this because some users choose to use standard price resources, some choose to use actual price.

You're right to do a good comparison across areas you would have to consider standard pricing. So I can tell you how we would do it if we were going to do it but that's not part of the submitted measure.

Bruce Steinwald: Right, so the submitted measure implies only that this would be used within a group or within a specific region as specified. Is that correct?

Dan Dunn: Or if someone went through the work of standard pricing data across regions, they could just supply the methodology as is.

Bruce Steinwald: Okay. Other questions around usability? Ashlie or (Sally) did we discuss usability satisfactorily enough that we will have covered or do we need to continue on this?

Ashlie Wilbon: Yes, just like a little bit, maybe more thoroughly, when you go through each one and just make sure that we have an understanding of what your - the consensus of the group is and where you stand.

So 3A looks at whether or not the measure performance results are reported to the public at large in national or community reporting programs.

Bruce Steinwald: All right so that's a very specific question, right?

Ashlie Wilbon: Right.

Bruce Steinwald: Has it been reported to the public anywhere that it's been used?

Ashlie Wilbon: Right, so I don't know if you guys have discussed that or...

Ashlie Wilbon: What I heard, and now, so and I guess maybe take away what we heard and then get reaction to help facilitate this. I heard that they are being reported in various and particular within organizations or within community situations that they are being used in that way.

Bruce Steinwald: Yes, but I didn't - I didn't hear from Ingenix whether they're aware that any of them had been reported publically to date.

Dan Dunn: Yes, Cheri, I know you did a survey to some extent of our customers. Do we know any one that are using this population base measure to do their reporting. I know a lot of them are still, you know, share information with physicians or even have it effect their, you know, incentive based programs. But is anyone doing sort of putting them on a Web site kind of thing?

Cheri Zielinski: Yes, I don't know if I'm at liberty to divulge the name.

Bruce Steinwald: Well, is it public or not?

Cheri Zielinski: Well, I - its' on a Web site that goes to their providers and then their providers get their risk scores and their scorecard and then they're able to drill down on that scorecard to the claim base level, the patient level, and then the claims level.

Bruce Steinwald: That's actually not the question. The question is are those publically available to the participants in the plan?

Dr. Tom Rosenthal: Or to the public.

Bruce Steinwald: To the public.

Ashlie Wilbon: Right and this is (Sally) and I don't if Barbra Rudolph is on the phone. I know that how to define the public at large is something that even the CSAC is grappling with. I think what we're hearing there's an example of it being reported it's perhaps not available to any person. You have to have some relationship. That's something for all of you to consider. At what level is it being reported?

I mean public at large I think most of us would agree at least from what I've heard in the conversations. That resource use measures are not necessarily targeting the consumer on the street. So some may disagree. Sounds like there are efforts to report it. You guys will have to weigh in on how much that meets your idea of what the public at large should be for these measures.

Female: ((inaudible)).

Bruce Steinwald: So the question is very specific....

Ashlie Wilbon: Yes. Barbara is that you?

Bruce Steinwald: ...actually as this available to the plan enrollees.

Ashlie Wilbon: That's not what the question necessarily.

Bruce Steinwald: That's my question.

Ashlie Wilbon: Yes, so Barbara Rudolph are you on the line?

Bruce Steinwald: I would like to hear it from Ingenix.

Dan Dunn: If there's nothing that prevents that from happening, I don't - I think Cheri's saying the same thing - do not know of anyone who. I do know, maybe to help you. I do know folks who use ETG based, episode based comparisons that have reported physician results to their membership. I do not know of a parallel use of that type of medium of communicating results for this population based measure.

Cheri Zielinski: Now I would add to that, this is Cheri, I would add to that. There's a lot of reason around that. Is an ERG score is one thing. Being able to drill down to understand how that score is derived, say claim level information, there's some PHI, you know, protection that needs to taken into consideration.

Bruce Steinwald: Cheri, I think we're talking about physician level or whatever. Whatever level results rather than the details. That was probably more the question.

Dan Dunn: I wouldn't see a problem in people doing that. I just do not know right now that that's being done.

Dr. Tom Rosenthal: Well, actually there's a problem because if you looked at - you made comments previously on how the data aggregated or scored. It didn't make any difference if it was a population based thing. If you took the melanoma and threw it in to the population base, our resource use calculation - all of a sudden somebody who's taking care of congestive heart failure - following your logic would get credit for the malignant melanoma.

So you're right, it doesn't make any difference for resource use or population based study. But your methodology sure makes a big difference if you try to use this on an individual level.

Dan Dunn: I'm sorry, I didn't get.

Dr. Tom Rosenthal: Never mind.

Dan Dunn: Yes, I guess I don't see the problem there but I'm sorry if I'm missing something.

Bruce Steinwald: Boy, if I interpret that comment, again, we had this discussion last time with the previous population measure that some things can skew this and create out liars. And again those out liars can be explained at the individual doctor level because you can drill down and go, oh, doctor X had a melanoma last quarter and therefore that explains why his cost index is off the chart. But when you're going to report that same number publically, you really don't have that opportunity to explain that.

And it gets back again to the potential difference between the use of this kind of information for performance improvement and even incentive pay versus the validity of this as a serious public reporting measure that's going to publically identify doctor X as being inefficient versus doctor Y being efficient.

Does that partly capture the sense of the comment in question?

Dan Dunn: No, I can see your point. No, I guess maybe my though is that melanoma would, you know, if the methodology's working as it intended would be included into the overall (inaudible) of the panel and should still let you do apples to apples.

But you're right the ability to understand, you know, below a designation or what's below a, you know, percentage above and below appears without knowing why is an issue. You're right.

Bruce Steinwald: And then I would ask (Sally) and Ashlie on this 3A, it's factual as to whether this measure specifically has been used for public reporting. However, we want to define public reporting. If the answer were no, that still doesn't mean that the group couldn't say but it could be perfectly acceptable for public reporting we just would have a kind of negative in the fact that it hasn't been used.

Ashlie Wilbon: That's correct, so.

Bruce Steinwald: Okay.

Ashlie Wilbon: I think 3B gets more into whether or not it's meaningful for the intended audience.

Bruce Steinwald: Which includes the public.

Ashlie Wilbon: Potentially, so that's something that's very important for the steering committee to also weigh in on. Who is the intended audience and, you know?

Bruce Steinwald: But I believe again we clarified last time. In every one of these measures the public is and has to be an intended audience or we wouldn't consider this to be an acceptable measure.

Ashlie Wilbon: Right, public reporting. But there are discussion about does public reporting mean that it's free, available on a Web site for anybody to Google in or is it within a network or is it publically reported in other avenues. So, you know, what the intended audience is, is an important part of the conversation.

Bruce Steinwald: All right well we got six minutes left, can we tackle this one?

Ashlie Wilbon: Almost, I think given the amount of time that we need to pause and do a public comment, wrap up, and I believe we're going to just have to put usability back on to the agenda for the person at least we got the conversation underway.

Bruce Steinwald: Very helpful.

Ashlie Wilbon: Hi, Allen; if you're there, could we queue up the audience or anyone on the public line for questions if they have any?

Operator: Yes, ma'am.

Ashlie Wilbon: Thank you.

Operator: The question answer session will be conducted electronically. If you would like to ask a question, please do so by pressing the star key followed by the digit 1 on your telephone key pad. If you're using a speakerphone, please make sure your mute function is turned off to allow your signal to reach our equipment. Once again, ladies and gentlemen, if you would like to ask a question over the phone lines, you may do so by pressing the star followed by the digit 1 key on your touch-tone phone. We'll pause a moment to assemble the queue.

Our first question comes from Kay Jewell from Centers for Consumers of Healthcare.

Kay Jewell: Hi, I just want to weigh in, it's not a question. The whole issue of who the public is, is extremely relevant if we're talking about measures that will be used by public payers. Then the public is anybody who will be influenced or affected by that measure and as a public consumer, it's not only the consumer having access. But it's also if the manager is going to influence the provider through the payers because of incentives or value based purchasing. Then it's relevant to the consumer to know what's influencing the provider that they may be selecting. That's it.

Bruce Steinwald: Thank you. Are there other questions or comments?

Operator: At this time, I'm showing no additional questions or comments.

Ashlie Wilbon: Okay, so it sounds like we've completed the public comment, we've got five minutes left, so. Let's just go ahead and wrap up for now. Just a couple of brief reminders. We are going to add the usability and feasibility discussion to the agenda for the in person meeting. And we'll

actually have you guys do your voting and final rating on site using the electronic voting tool that we have. And we'll go over that when you here in person.

If you have not already done so, hopefully everyone has arranged their travel for those of who will be able to make it to the in person meeting next week. And we, again, will send out a reminder for the final voting on the HealthPartners measure today for those of who are not - who did not submit your final vote.

We'll be seeing everyone next week for the in person meeting. And again, as always, feel free to email me or Sally or Sarah is you have any questions or concerns and I wanted to thank everyone and our co-chairs for a really interesting discussion today.

Dr. Tom Rosenthal: But we're still taking up Ingenix a little bit at the beginning of the meeting, right?

Bruce Steinwald: Oh, yes.

Ashlie Wilbon: I'm sorry?

Dr. Tom Rosenthal: We're still taking up this one at the beginning of the meeting.

Ashlie Wilbon: Yes.

Dr. Tom Rosenthal: And the voting is not open yet?

Ashlie Wilbon: No, no, we're going to - we're not going to send you a link for this measure online. We're going to have guys do it in person when you come.

Bruce Steinwald: Okay.

Ashlie Wilbon: Okay, thanks everyone.

Bruce Steinwald: Thank you, see you next week.

Operator: That does conclude today's conference, you may now disconnect.

END