

NATIONAL QUALITY FORUM

Moderator: Ashlie Wilbon
June 6, 2011
1:00pm CT

Operator: Welcome to the conference please note today's call is being recorded. Please stand by.

Ashlie Wilbon: Hi, good afternoon everyone. This is Ashlie from NQF. I'm joined by Sally Turbyville and (Sarah Fanta) and (Moran). I just want to thank everyone for joining us today.

We are going to try to jump right in to finishing up the evaluation of the HealthPartners Measure as well as starting the Importance discussion for the Ingenix Measure.

So we do have - we should have hopefully both developers on the call by the time we begin discussion. We're just going to do a brief roll call. And to see who's with us today and then we'll jump right in.

Doris, are you there?

Bruce Steinwald: Bruce is here.

Ashlie Wilbon: Bruce, okay. Paul Barnett?

Paul Barnett: Yes, I'm on.

Ashlie Wilbon: Jack Bowhan, I know you're there. Jephtha Curtis? (Curtis Elword)? William Golden? Lisa Grabert?

Lisa Grabert: Here.

Ashlie Wilbon: Ethan Halm? Okay. Ann Hendrich? Tom Lee? Jack Needleman?

Jack Needleman: Here.

Ashlie Wilbon: Okay, hi Jack. Doris Peter? Mary Kay O'Neill? David Penson? (Steve Phillip)? David Redfearn?

David Redfearn: Good morning.

Ashlie Wilbon: Hi there. Bill Rich?

William Rich: Yes.

Ashlie Wilbon: (Jeff Rich)? Tom Rosenthal? Barb Rudolph? Joe Stephanski?

Joe Stephanski: I'm here.

Ashlie Wilbon: Okay. (Jim Weinstein)? And (Delores Shanagahara)? Okay. Small group today. All right, so let's go ahead and get started. Again, our objectives today are to complete the discussion of the HealthPartners Measure and to hopefully wrap up with a discussion of the Importance Criteria and for the Ingenix Measure.

So I won't go into the details about the criteria again but Bruce maybe I can hand it over to you here and you can decide as a group whether or not you'd like to review some of the scientific acceptability analysis that was submitted by our statistical consultant Carlos Alzola, whose actually on the phone as well to answer any questions that the Steering Committee may have about what was said.

Also, the HealthPartners Developers, Chad Heim and Sue Knudson are on the phone to answer any questions you may have about that as well. So with that said, Bruce connect with you.

Bruce Steinwald: Well, I certainly think it's a great idea since we have two new sources of information to take advantage of those sources. And I guess I'd open it up to the Steering Committee to raise any questions or issues of either one of those sources at this time relating to the matter of scientific acceptability.

Male: I'm sorry, what are the two sources. I thought we had the statistical council. Is there something else I'm missing?

Bruce Steinwald: Well, we also have people - the person or people from Ingenix, I'm sorry, from HealthPartners...

Male: I see.

Bruce Steinwald: ...on the line as well.

Barbara Rudolph: Hi, this is Barb Rudolph.

Ashlie Wilbon: Hi Barb.

Thomas Lee: And Tom Lee has joined too. I missed the very first few minutes but I'm here now.

Ashlie Wilbon: Okay, thanks.

Bruce Steinwald: That was just the roll call, anyway Tom. All right, this is Bruce.

Ashlie Wilbon: Should I...

Bruce Steinwald: Yes.

Ashlie Wilbon: Should I bring it up on the screen. I'm not sure people...

Bruce Steinwald: Yes.

Ashlie Wilbon: For those of you who are in the webinar I do have -- I can bring it up on this screen here.

But I realize it takes a couple of seconds for your screen sometimes to catch up. So bear with me here.

But what our Statistical Consultant completes for us is essentially a worksheet, a standardized worksheet that we've developed with some questions on there to give you an idea of the appropriateness or soundness of the reliability testing, the validity testing and the risk adjustment model. So I will use it...

Bruce Steinwald: Well, unless people want to raise specific questions we might have the consultant, would you be willing to kind of summarize your findings on those three areas, reliability, validity testing and risk adjustment?

Carlos Alzola: Certainly. Does anybody have any specific questions?

Ashlie Wilbon: Maybe just a summary Carlos. Or...

Carlos Alzola: Okay.

William Rich: ((inaudible)) -- this is Bill Rich, Carlos and it looks like you had some concerns about the risk adjustment on both of them and I'd like you to address that.

Carlos Alzola: Okay. Okay. Let me start then with reliability. We look at reliability as -- we look at was the data reproducible. We look at relating the sense of was the data reproducible.

As we look at where you see claims data really the only thing about reproducibility means whether the, at the level of the closing we can reproduce that.

And since it is - we are using commercial database we can, that should have been done at the level of the closing, by the database mantle. However, the - in this case they did what I saw was a very good job in terms of examining what is the availability of the data, what is the significant measure.

One thing that they did was to do a sampling, two types of sampling. In one case they selected a 90% sample without replacement from provider group and compared the reliability of the sample to the actual bodies.

When they say we start resampling that means that they are - they select one patient at a time from the database until they complete 90% of the data. So that would give us some idea of what the influence of extreme (values) may be.

Not all the time the extreme bodies would be select. So they select 90% of the data multiple times, there's with 500 times and then they compare the results that they obtained from the other ratio for those 500 samples to the - back to the results from the whole sample.

And we found that there's very small change. It could - the difference between the actual -- and let me focus on the total cost index -- the difference between for each provider, the difference between the samples and the actual TCI range from negative .0069 to .00083 which is in percentage since we are looking at values that range from about .8 to 1.1 is a very small percentage so that's only .1%.

So that's by itself tells that the influence potential influence of these extreme values would be very small. The other approach to use was a (width strapping) which is very similar except that instead of selecting a 90% sample without replacement. They select a sample of width replacement so the same person could have been the sample twice.

So with this test it stimulates the reliability that you can find in the population very effectually. This isn't necessarily very widespread right now, nowadays and so there simulating reliability in the population and so they found very, very small ranges of change between a sample and a potential difference sample from the sample population.

And what they actually are saying. So that show again very small reliability with respect to the actual values itself.

Paul Barnett: So is this the overall -- this is Paul Barnett -- this is the overall variability of the -- so for all providers?

Carlos Alzola: They do this at the level of the provider and then you look at each provider and compare the difference between the (width) strap samples and that provider and so all of the differences range from a very small .000067 to .00252.

Paul Barnett: I was just wondering if you're familiar - there was a paper that was given to the Steering Committee at the outset of this on the reliability of one of the episode groupers.

Carlos Alzola: Yes.

Paul Barnett: The first author was Adams.

Carlos Alzola: Yes.

Paul Barnett: And they developed a method of signal to noise where they as I understand it looked at the between physician variance compared to total variance. And then they had a way of expressing reliability on a scale of 0 to 1 and wondering if that's - so that seems to be another way of looking at reliability.

Carlos Alzola: Yes. I haven't seen that paper.

Bruce Steinwald: I think - well I think that reliability on an individual provider level.

Paul Barnett: That's right, so that's what I was understanding what - this was purported to represent.

Carlos Alzola: And it does because we're looking at - they're looking 18 divisional providers and they are stimulating the reliability that they would have served in the population of that provider. So by selecting a sample from that provider 500 samples actually and allowing them to select the same patient more than once.

What is simulating is the reliability that you would find in the population for that provider. And when compared to the actual value that was observed we find that, you know, for our provider the difference was very small.

Bruce Steinwald: Any other questions? Go ahead.

Paul Barnett: Well, I'm just saying that it would be, you know, there seems like a different method than what Adams used. And I'm trying to reconcile the two. So they separated the components, variability scores among physicians and variability scores for individual physicians.

So they sort of decompose the variance and...

Barbara Rudolph: This is Barb Rudolph. I think, yes, what you're saying in this one that it's by provider group not by an individual provider.

Paul Barnett: But even if you define the provider differently it seems like you could still have that same sort of measure of reliability. David perhaps that would be a useful thing to consider here.

It's just of how these two methods of reconcile or don't reconcile that.

Ashlie Wilbon: So Paul this is Ashlie from NQF, we don't actually require that the developers use specific methodology to demonstrate reliability. But in the method that they do choose they should be able to demonstrate that the measure is reliable.

And, you know, justified by the results that that has been done. So we don't - I realize that there was a difference methodology used in the paper but we don't - we don't necessarily require them to use a specific methodology. So...

Paul Barnett: No I understand that, my concern is, is that the findings are sensitive to the method.

Male: The other is I think they're measuring two different things. One is the and I'm trying to go back to remember that paper but I believe that was on the individual provider level.

Paul Barnett: Right, but whether you define the provider as an individual physicians or a group of physicians it doesn't really matter. The question is, is there - are we measuring signal or is it noise that we're - that we're getting.

Is there - and that was the notion in the Adams paper that somehow how much of it's just natural variance and how much of it is actually is something that is important that we care about, consistent, reliable.

Carlos Alzola: Well, I think that what this is shown is that the noise is small relative to a signal.

Chad Heim: This is Chad Heim, Carlos I think part of the issue, the Adams thing was looking at how reliable, how precise the individual physician reliabilities. When - just - I want to make sure we're talking about the same thing.

At what level is your reliability the analysis being conducted what are you trying to measure with the reliability?

Carlos Alzola: This is performed at the provider level.

Chad Heim: So it's at the provider level?

Carlos Alzola: Right.

Chad Heim: Okay, so we're talking about the same thing in those two different analysis?

Bruce Steinwald: Sounds like it.

Chad Heim: Okay.

Carlos Alzola: Yes, and again I'm not familiar with the method that you're discussing. Presumably it can be used at the provider level as well. But as a matter of looking at signal to noise it's just to - you just want to see how much, reliability there is at the provider level between what was observed and what could potentially had been observed.

And the way that - one way to do that is just by simulating the variability that you would observe in the population and the (width strap) is being in use for a few years now and it's being - shown to be very effective.

Paul Barnett: Well sure, it's a good way to measure - to estimate the variance and so what we're estimating the variance of is the observed to expect it to ratio?

Carlos Alzola: Yes, we're looking at the variance of the observed to expect to ratio, that's correct.

Paul Barnett: All right.

Bruce Steinwald: Okay, can we - can we move on to validity testing?

Carlos Alzola: Yes. Validity, for validity they compare the values that they would obtain - the values that they obtain in terms of the risk adjusted values and the no-risk adjusted values to some other indicators for resource use.

One of them was compared to total, total care relative versus values that was the basic units for measuring resource use. And they compared to the actual, per member per month spent.

So we spend those values would have a hike of relation cost because most of use would have a higher call and so we really - we found a hike of relation there of .98.

Now those - that was the comparison of the non-risk adjusted values to the actual money spent. After the risk adjustment was applied risk of relation went down to really 2.15.

So that would show that risk adjustment is doing a good job of reducing the dependency between those two values. It's not perfect. I would have expected it to go down a little more, close to zero, but still it's a good - it does a good job of reducing - of adjusting for techniques.

Another example is when they show the correlations, restricted to the different places of service. So they look at what the correction was between the total resource used to the risk adjusted values.

For example for the risk adjusted patient services and then again the correlation was very high. And so there were a number of this kind of test that were performed and they show variation of the correlation was what was expected.

It was high in most cases and most importantly for my - in my view was when the risk adjustment was applied then relation was - was lower as it would be expected, showing that the risk adjustment is doing its job.

Bruce Steinwald: Carlos you were talking about both validity testing and risk adjustment at the same time? Was there something further you wanted to say about risk adjustment?

Carlos Alzola: Risk adjustment the only thing I can say about risk adjustment is that it was a methodology that it was accepted by the Society of Actuaries and then they did - the Society of Actuaries did a photo analysis of several risk adjustment methodology.

Bruce Steinwald: Okay.

Carlos Alzola: And this one was one that was validated as best being - having a type of relation with resource use. Now I don't have any details on how that - how the risk adjustment is actually performed since I really don't have access to the - to the details of the methodology.

All I can really say is that it's being approved by this organization. And the way data that Society Actuaries submitted suppose the claim that it's really an effective risk adjustment methodology.

Bruce Steinwald: Okay.

William Rich: Carlos, this is Bill Rich is the fact that it's not available because it's proprietary or we just didn't have a password to get into the Hopkins site?

Carlos Alzola: I think I need a User ID and a Password. Yes.

William Rich: I noticed that certainly I couldn't get on. I think Barbara had noted some comments that some other people did.

Carlos Alzola: Right.

William Rich: Do we know if it's proprietary, I don't understand why we couldn't get access.

Ashlie Wilbon: Maybe Sue or Chad could comment on that.

Sue Knudson: Hi, sure Ashlie this is Sue. Hopkins software is a proprietary software that you have to license. So that could be part of it but one update that we had since our last discussion a few weeks ago was they did just recently announce the software free of charge to the health insurance exchanges and to plans under contract with the exchanges.

My guess is that it's still in development from that announcement. I am remote for the other Steering Committee members to know. So my colleague Chad is so would just advise him to comment further if he has additional to add.

Chad Heim: No further comment.

Bruce Steinwald: But for the ordinary user, the software is available for a fee right? But there's no, no other requirement or do they impose requirements on users that they, they be of any particular type nonprofit organizations or anything like that?

Sue Knudson: No, there is a licensing fee and there's a scale for the size of the organization and the type of organization.

Bruce Steinwald: Yes.

Sue Knudson: And we did provide some information from Hopkins to build a supplement on - just some basic information about what that structure is.

Paul Barnett: Yes, \$33,000 plus 2 cents to 28 cents per member.

Bruce Steinwald: Yes, so it's not cheap.

William Rich: ((inaudible)) and we don't even have a way to understand the risk adjustment, use administrative claims data is it all group or software. I'm really very bothered by the fact that we're not able to make a value judgment on one of the key things that we're supposed to look at.

Sue Knudson: So as a part of our application we had submitted the risk adjustment under the guideline portion of the measure indicating as Carlos had reiterated that we're really relying on the Society of Actuaries finding in terms of what is the most robust risk adjustment to apply to this type of measure to make sure that the end result is indeed, you know, credible and usable.

And so that's why we applied with the flexibility of using any of those tools. The one that's not on that list is HTCs which is the open source tool from CMS and I think that's only because the study predated the creation of that tool.

Bruce Steinwald: Yes.

Sue Knudson: But -- so that one would need reliability and validity testing that was our recommendation if another user was to use that.

William Rich: Can you even tell us are they using administrative claims data, within the administrative claims data, are you free to even discuss the generalities of the methodology?

Sue Knudson: Oh yes, I think that's very common knowledge because it's looking at eligibility to the degree that when you apply the eligibility there's a continuous enrollment criteria and then claims data to look at diagnosis.

And age and gender as well but that can be obtained from eligibility.

David Redfearn: This is David Redfearn. This is a kind of a funny situation which I don't exactly understand. Ingenix which is certainly very proprietary oriented in terms of their methodologies.

They have completely opened up all the methodology issues for the ERGs ETGs and all those methods. They have a website, a transparency website in which you can get access to the complete technical documentation of how the models work, how you run them and everything.

And obviously you can't run the model if you don't have a license for the software but all those details are available. It seems so little surprising that something equivalent to that is not available for the ACC.

William Rich: David you just took the words right out of my mouth. And all - everybody that's applied too for the - all four applicants to CMS group or contract. Even before that you made these things probably available on their website so that you could drill down with any, any database to check the reliability and the assumptions made on risk adjustment.

So I like you, find this kind of troubling that we're unable to address it. Since we can do it with the four, you know, ((inaudible)), Thompson Reuters and Ingenix they're all out there now in the public domain.

Bruce Steinwald: Yes. What does Hopkins make available about it's methodology virtually nothing?

Sue Knudson: And if I can just clarify, you know, as I said, we're indicating that, you know, we just want our suggesting through guideline that any risk adjuster could be a paid and use that was endorsed by the Society of Actuary study.

So to the extent others may have greater levels of transparency, that could certainly come into play here and they could be used with this, with this measurement approach.

So we just have a history and they're using the DFT tool which sets the vendor for that ACG software. So that - that was really the guideline portion of the application.

Paul Barnett: So one question that's kind of corollary to this -- this is Paul again -- is that initial part of this that says, you know, does the measure steward have the intellectual property rights to the measure.

And I was a little bit concerned that, you know, in order to use this you have to have a measure that is, you know, the measure steward doesn't actually control.

And it's true there are alternatives but all of the validity test and everything are used one particular case mix measure. So we're sort of buying into that one case mix measure if we endorse this particular submission.

And I'm concerned that it doesn't meet the measure stewards in the criteria that the measure steward owns the property rights to measure it.

Sue Knudson: So this if I could just connect one last dot, that is where when Carlos was mentioning the society of Actuaries findings what they found in the most basic sense was all of those commercial risk adjusters that they were evaluating all essentially had equal or comparative levels of predictive power.

And explanatory power. So they're working and operating at the same level.

Mary Kay O'Neill: This is Mary Kay O'Neill and I just actuarial standpoint they're equivalent that's one thing but does the use of different tools change specific outcome of the application of the measures.

Paul Barnett: In other words did the providers get rated the same way?

Mary Kay O'Neill: Correct.

William Rich: Well, I think that's a very key question, and this is Bill again Mary Kay, if you look at the, if they do, how should the same level of the other three commercial software products. We do know that they have been rejected for people over 65, Medicare population ((inaudible)).

So I guess we can't certainly for complex patients, patients of multiple care mobility's all of those types have been rejected by Medicare.

Mary Kay O'Neill: Okay.

Chad Heim: This is Chad, just to kind of expand a little bit on what Sue was saying now when we initially submitted this we weren't endorsing a particular risk adjustment product. So that's why we were referring back to relying on a site that actually went an evaluated all of these groupers.

And so that's why we're just suggesting this as a guideline versus testing all the groupers to this measure submission. And in the Actuary, the performance is similar so if you can translate that into the final TCI scores would be similar as well.

They just all go about it slightly different but the end product is similar.

Bruce Steinwald: So are you saying that from, from HealthPartners standpoint the measure could be applied with any one of these risk adjustment methodologies?

Chad Heim: Exactly yes, and have similar results.

Barbara Rudolph: This is Barb Rudolph. A couple of things I'm just thinking that I don't think the measure developers, you know, at this point are required to test their measures using multiple, you know, different proprietary systems or non-proprietary systems.

And I guess the other thing I just want to mention is there's hundreds of peer review journal articles on various specific topics that have utilized the ACG system and I have no proprietary interest in this at all.

I just think that there are some things that we're not going to know as a Steering Committee and that is exactly how these various prices work. So we know that, you know, they sort of help up to the, you know, test of peer review or whatever you think of that.

And that we're seeing some of the documentation for these various systems likes this it would take you a very, very long time to review and really understand how these - how these particular programs work under hundreds of pages of documentation.

So I don't know. No, I myself don't want to go through 300 pages of documentation. I feel like I need to trust someone else's judgment. So, you know, suggesting that perhaps we can accept this without knowing exactly how it works. That may be radical.

Bruce Steinwald: Part of your argument is that it's been used many, many, many times and it's -- the work that's been done using this risk adjustment methodology has been published and peer reviewed.

Barbara Rudolph: Right.

Tom Rosenthal: This is Tom and I would agree with that point and in fact this is not a black box. It's available, I mean we could get it. It would require a little effort but we could get it unlike some of them which are completely proprietary and therefore a black box.

So I think there is a differentiation here. It is worth knowing though I think in some level how discriminating the risk adjusting methodology is because if you look at the spread between the worse group and the best group in the analysis it was plus and minus 5% from the mean.

And the question in regard to signal to noise ratio I would suspect that there's at least a 10% noise that's left inherent in however good the Hopkins risk adjustment methodology is that may or may not make this thing terribly discriminating.

But I'm not sure if that's our problem in terms of deciding whether or not to endorse it.

Bruce Steinwald: Any further discussion?

Mary Kay O'Neill: This is Mary Kay and my understanding of risk adjustment methodology in general is that they only account for at very best less than 50% of variation. So I mean, you know, people put a certain amount of credence in them and they can only tell you so much.

But my point is more that if we're endorsing a measure it's going to be broadly utilized, we just need to know how they'll work in serious system. And I don't care about knowing the whole standard page detail analysis of the data methodology.

So I just - I just need to know that if different utilizers of a measure such as this pose different risk adjustment methodology, the results that they would be getting would be really comparable to results to some other systems and services and different ones.

And that's my only real question.

Bruce Steinwald: I believe the HealthPartners representatives said that they - he believed that they would. But I think that's based again on the Society of Actuaries not their own work.

Mary Kay O'Neill: And to tell the Actuaries were saying that the different methods has the same power that's fine but, you know, power and behavior of methods isn't exactly the same I don't think. But anyway that will be -- I'll stop there.

Jack Needleman: This is Jack Needleman. Can I ask the HealthPartners person a question?

Sue Knudson: Sure.

Chad Heim: Sure.

Jack Needleman: Yes, there are different risk adjustment methods. Can you say a little bit about why you chose this one?

Sue Knudson: We actually have a long history with John Hopkins going back to, you know, the development and the initial data testing of this tool. What I can tell you from, you know, just an operating health plan working with several different employer groups who might have different insurance carriers across the country.

We're fairly familiar with working with the different one's as well and smaller scales. So, you know, that's basically why we have stuck with ACG is because we have that long history with them.

The other piece on ACG is it is a diagnosis based risk adjustment tool which isn't relying on delivering units of services in utilization. To come up with its methodology and we tend to like that diagnosis based method over a utilization based method.

Chad Heim: In addition, this is Chad from HealthPartners - in addition I'd add too that this market area here department of health actually uses the ACG grouper as well. So in this market it's currently ACG. But like Sue was commenting we've used ARG, and CRGs and depending on the application.

Sue Knudson: Yes, it's somewhat communities standard. There's a lot of use in our market.

Bruce Steinwald: Any further discussion?

Tom Rosenthal: Can I -- I'm sorry -- Tom Rosenthal again - I wanted to quickly ask a question about the attribution. I know that was covered earlier but I missed a little part of it. If I understand the rule is that more than 50% of the visits to a primary care physician is what sticks that patient to that provider.

Do you have data though on what percentage of the patients say had more than 40% of their primary care visits with a minority provider?

Sue Knudson: Yes, so attribution we actually go on the greatest percentage of visits.

Tom Rosenthal: Right, so that would be 51%.

Sue Knudson: Well, the greatest percentage is not a majority.

Paul Barnett: Yes, if they had five per rider it could be 21% right.

Tom Rosenthal: Got is, okay, well I misinterpreted that. So what percentage of the time was there, you know, more than one provider, primary care provider?

Sue Knudson: Yes, we actually - so we're attributing at the group level and we found that that model we have -- I think it's about 10%, is that correct Chad? That we're not able to - that go to more than one provider?

Tom Rosenthal: You mean at the group level?

Sue Knudson: Correct.

Tom Rosenthal: Well, that's why somebody made the comment earlier that it really doesn't make any difference whether it's the individual or the group and I think that's not exactly true.

Sue Knudson: Well, let me clarify.

Tom Rosenthal: The groupness washes out an awful lot of the attribution issue.

Sue Knudson: Well, for the number of members but then also we apply a threshold for...

Tom Rosenthal: 600.

Sue Knudson: Yes, 600 and so we had just based on the Steering Committee discussion at the last session, we had done some additional testing just knowing based on that discussion that the - that the incises might get smaller because other markets are group practice bases.

Tom Rosenthal: Right.

Sue Knudson: So we had done some additional reliability testing and just had prepared a slide to share that with you all. So you can make that translation out of our more dominant group practice market which might be different from the rest of the country.

Tom Rosenthal: But I think this is a relative point because in the rest of the country the misattribution or somebody has patient to see five primary care providers and yet the one who saw 21% of the visits is attributed all the cost is I think problematic.

And I do think in a group business practice model you wash out a lot of that variation because patient may see two doctors within the same practice but as you say only 10% are going from practice to practice.

Thomas Lee: You know, this is Tom Lee, just using analogy products in ECG and so in our network, you know, those theoretical problems are, you know, we understand them. We find that it doesn't really emerge as important in their actual application. Like there doesn't seem to be systematic error that's introduced.

The main issues is like selling it to the doctors who are being profiled. But if we raise that percentage and require a majority of claims for example then the number of claims is going to attribute to any doctor goes way down.

Tom Rosenthal: But I do think I agree Tom that that point is well made. But it's seems to me selling this to the provider groups is an important aspect of this if it's going to be used to try to drive change which I assume is substantial reason for the development of these groupers and their application.

Thomas Lee: No, I hear you at the end of the day we ended up just settling on 30% as our - our threshold. So it's possible for an episode to be attributed to three different doctors or no doctor.

Tom Rosenthal: Right. I guess it's one thing when it's used for internal quality improvement but this is the same issues that we have with many of the quality measures. If you're using it for quality improvement and your doctor largely accept it when you produce in your health plan the information from them and they decide that they want to work on it.

Everybody is happy the instant it's starts getting published in the New York Times is a national measure and somebody's got a big asterisk by their name because they're appear to be five percent more costly than everybody else and there may be other significant financial consequences to that.

The acceptance at the provider level becomes a nontrivial problem it seems to me.

Thomas Lee: No, I think that's a very good point it gives them an opening to criticize it and, you know, of course I haven't seen anyone on our system who would admit to being happy.

So I wouldn't go that far but I think you're points well taken.

William Rich: I have a question about the attribution. If somebody had an expensive hospital stay, but never had a primary care doctor and then they picked up that primary, then they got one after discharge and it was after their discharge, that primary care provider would then be assigned the cost of that hospital stay, right? In their profile according the attribution rule?

Sue Knudson: Well, you know, it depends how the study period is defined. Let's say if all of that happened in a calendar and their primary care visit within December. I mean conceivably that could - that could happen.

William Rich: Then my only concern would be then there would be a kind of disincentive for someone to pick a patient like that.

David Redfearn: This is David Redfearn. There's a related issue that we see a lot in California. And that is when you - when you build these attribution models around selecting PCPs we have a lot of care that seems to be managed by specialists that the patient never really sees the PCPS.

And when you kind of force these, these attributions to select a PCP, you get these awkward situations in which you have a PCP that you selected but really some surgeon or something is driving a lot of the cost to the care.

One of the things we have done in terms of our attribution model is that we look - we do the attribution differently based on whether the episode of care is a medical or a surgical episode.

The medical ones we try to use contact measures that's in -- and basically that's going to often times get you a PCP. But for surgical episodes lots of times we look at specialist and you get the surgeons and the surgeons are driving a lot of cost and so that can - does a make a difference.

I do want to kind of second the comment made earlier is that in a practical sense also oftentimes this doesn't make a heck of a lot of difference. Looking at individual physician profiling somewhere around 60% to 65% of the episodes only have a single physician involved in the episode.

So it's not ambiguous then at all. The other comment I'd make is when you do push this up above the individual physician and look at a group in our case it's like Tax IDs when you have these multi-specialty medical groups that removes these kind of issues.

Because even if you have a specialist managing care often times there in the same medical group. So you're going to assign it to the same medical group. So again this is very complicated and if you look around the edges these things make a difference and a big picture often times doesn't.

Bruce Steinwald: Okay, this is Bruce again. Any further comments on scientific acceptability issues?

William Rich: Yes, this is Bill again. Again, I'm disappointed we were hoping to see more transparency.

Bruce Steinwald: Yes.

William Rich: In all three areas and our first crack at risk adjustment and it's actually cloudier than commercial products that are out there. And I'm sorry to see that.

Bruce Steinwald: Point taken. It does seem to me though that we - we are faced with a situation here either finding an acceptable or not. I don't think that we can - we can ask HealthPartners to change it, so risk adjustment methodology at this point.

So I guess we're then like them either reliant on the Society of Actuaries analysis and on the fact that the method has been used many times and published in peer review articles. Is that sufficient for us going forward or not?

I think we just have to make the judgment.

Ashlie Wilbon: Hi Bruce, this is Ashlie, just wanted - we just wanted to add a note typically on the discussion with the concerns around the level of analysis at the physicians versus the individual physician level versus the group level that they have actually specified the measure to be used at the group level.

So in terms of you know users out there using the measure if it's endorsed it would be endorsed for use at the group level. We're hoping that helps a little bit.

David Penson: This is Dave Penson. I've been sort of lurking listening. I have to say that the fact that it's working at the group level as opposed to the individual provider level is really what swayed me with this.

I didn't have the problem with risk adjustment as everyone else seems to have. Based simply on what Bruce said, I mean look it's not a perfect system but it's in the peer review literature and, you know, I don't know a better way than to do it.

So I personally think it's acceptable and I don't know if we need to take a vote or what but I think we're at that point, either you buy or you don't.

Barbara Rudolph: This is Barb Rudolph. I agree and, you know, I guess I up to vote to me, on mine so...

Bruce Steinwald: Ashlie what is required of this Steering Committee at this point, is it sufficient for us after this discussion simply to complete filling out the worksheets?

Ashlie Wilbon: Yes, what we're going to do actually -- at this time we'll probably have you guys move on to usability and feasibility...

Bruce Steinwald: Right.

Ashlie Wilbon: ...discussing those criteria and then at the end of the call actually we're going to send an email out to everyone with the link for you to evaluate the measure with your final ratings and your recommendation for endorsement.

So, based hopefully, you know, everyone has heard what they need to hear in order to make their individual ratings in voting and vote by the end of the call.

Bruce Steinwald: Okay.

Barbara Rudolph: Excuse Ashlie, can you verify something for me?

Ashlie Wilbon: Sure.

Barbara Rudolph: On the ratings situation -- I'll give you an example. The possibility for errors in the data or something like that, it's one of the vote areas. If you say high on that, you select high. Does that mean that the date is high in errors or you have high reliab- you feel like there's -- it's a - it's a good data source and doesn't have high.

You know, isn't floored with errors?

Ashlie Wilbon: Right, so we can send clarification on that in the email. I believe - are you talking about the one of the feasibility criteria?

Barbara Rudolph: Yes, I think it was because it was...

Ashlie Wilbon: 4C.

Barbara Rudolph: Going in different directions and I wasn't sure how to score it.

Bruce Steinwald: I thought that high was always good. Is that true? High is always good, low is always bad?

Ashlie Wilbon: Right.

Barbara Rudolph: Okay.

Ashlie Wilbon: So I think you're referring to 4C? Where it's asking about acceptability to an accuracies errors or unintended consequences related to measurement? So if you rated the measure high then you're saying that there is low susceptibility to inaccuracies.

Barbara Rudolph: Okay, then I did it correctly. All right.

Ashlie Wilbon: Okay. Does that make sense to everyone?

Bruce Steinwald: yes.

Ashlie Wilbon: Okay.

Joe Stephanski: This is Joe Stephanski. I have one question here because it's become a little muddy along the way I think. When we rate this particular measure, this time through are we talking about rating it using the John Hopkins risk adjustment or are we literally saying that it would be acceptable to use any of those within the Society of Actuaries study?

Ashlie Wilbon: Heidi if you're there, I think I'm going to answer this and then she can correct me if I'm wrong. Because the measure has been specified and tested using the ACG John Hopkins methodology, that is what you would be rating the measure on.

Is that correct Heidi, even though they've listed other risk adjustment methods that can be used but because they tested it and all of their reliability is listed validity data is based on that particular model?

Heidi Bossley: Yes, this is Heidi. I think we had an internal conversation to staff on this one two because I think it's a balancing act that you all need to as a Steering Committee decide they've shown through at least the actuarial information that it's probably equivalent.

It appears several other types of risk adjustment strategies, but I would probably for the purposes of what you're doing right now it's how they are specified is how they'll use it. I would use it based on the ratings, rating it based on the ACG, risk adjustment put forward.

Joe Stephanski: Okay so we...

Heidi Bossley: If you could comment and you think that it can be applicable across others as well or any other information that would be useful to help guide us to write up how you came to the rating you did that would be good.

Joe Stephanski: All right because there are certain things about actuarial science that I have some issues with. And I do not see the model as equivalent or, you know, no matter what the Society of Actuaries may have said.

But I'm willing to rate it just on the use for the Johns Hopkins.

Ashlie Wilbon: And I think if you all could just provide us some information on how you did rate it, if it was just based on that or if you additional thoughts on it that would be helpful.

Joe Stephanski: Okay. Thank you.

Ashlie Wilbon: So if everyone's okay with moving on to the usability and feasibility, we would like to try to wrap up this measure. And then get to the importance of the Ingenic measure before we close.

So I'll pull up the usability criteria so you guys can have a reference.

Bruce Steinwald: While she's doing that is there anybody that would like to ask a question or make a comment?

Joe Stephanski: This Joe Stephanski again. Just another question about, this has been restricted to the commercial under 65 population. Is there something about any of this methodology that would prevent it from being used with higher age groups that those were, if we had databases available? In other words, for a Medicare population or something very specific here that I'm not seeing?

Bruce Steinwald: Well, I think the key is that the raw material for the measure comes from an employed based population.

Joe Stephanski: Right.

Bruce Steinwald: You know, so you're asking what if it was Medicare's claims data presumably you would come out with different waiting factors. But you're asking, well, why not. Why couldn't they?

Joe Stephanski: Yes.

Bruce Steinwald: I don't have any answer. I don't see any obvious reason why not. But maybe I'm missing something.

Sue Knudson: Bruce, this is Sue, if I could clarify. We're largely a Medicare cost plan. So we simply don't have a full claims data set. We're lacking hospitals. So we weren't able to run it that way. But I'd agree with what you're saying. Theoretically if you have that full claims database, you should be able to use it.

Bruce Steinwald: Okay.

Male: Sue, I was looking at your ((inaudible)) graph and making that assumption. Sue, if you claim that the answer the same efficacy of Cave, Ingenix, and Thomson Reuters they've already been rejected in the ((inaudible)) population by Medicare. I'm very uncomfortable with addressing that issue.

Sally Turbyville: This is Sally, just add to your conversation, this measure is specified for the commercial population only and also excludes patients that are over 64 for measurement. So if that can help you in your conversation.

Joe Stephanski: I'm just thinking in terms of the CMS value based purchasing and their reach for a per member per month kind of cost measure. And we're not going to have anything to offer them because we haven't been - and none have been submitted to us.

Bruce Steinwald: What I see on my screen is really, really garbled.

Ashlie Wilbon: Hi, I just changed the document; hopefully it will clear up in a second. I know sometimes it takes a second for the screen to catch up. If it's still garbled, what I pulled up was that side-by-side table that has each of the sub-criteria, delineated with - next to that the submission item that corresponds with that.

And I can read that aloud if needed, if you guys still can't see it.

Bruce Steinwald: Well, my screen is looking bad, what about other people.

Male: Mine's okay.

Joe Stephanski: Mine's okay.

Sue Knudson: Did you try a refresh, Bruce on your browser?

Bruce Steinwald: I'll do that. And in the meantime, let's forge ahead.

Ashlie Wilbon: Okay, great, for usability the first sub-criteria, 3A. Is as a measure performance results are reported to the public at large and national community reporting programs.

By the time of endorsement, exceptions may be considered if there's evidence that the measure performance results are available for public reporting and that use of the measure has benefitted the public.

Doris Peter: Hi, this is Doris Peter from Consumers. In the application there was a mention that data were publically reported but I couldn't find it on the Web site. And none of the attachments seemed to be what would be publically reported so I was just looking for some clarification on that.

Male: I couldn't find it either.

Male: Planning to.

Ashlie Wilbon: Does Sue want to comment on that?

Sue Knudson: Sure, I'll give it a start and let Chad Heim as well.

We use it on - in benefit design and in transparency on our healthpartners.com Web site. And so that's where we have been using it as a plan around transparency. We have some beginnings of a community collaborative as well but that's largely out where that work has taken place.

The other piece, there was a publication by our health partners Research Foundation with diabetes care just using it in a condition specific example. But understanding the inputs for resources to derive quality levels. And that was published as well.

I'm not sure if we included that as a citation. Chad what would you add?

Chad Heim: What - this we added on healthpartners.com we translated the affordability matrix (TCI) used to consumers speak. We're translating that into dollar signs. We've gone through focus groups putting index numbers out there, index numbers aren't as meaningful. So we translate those into dollar signs. But the underlying methodology's what we submitted.

Doris Peter: Okay, then I saw that on your site, I'll take another look at it. I thought that dollar sign was something different but I'll look again, thank you.

Ashlie Wilbon: Okay I'll just go ahead and read 3B aloud for those of you who can't read the screen. The measure performance results are considered meaningful, understandable, and useful for the audience for both public reporting and informing quality improvement. This can be supported by rational - a rationale or demonstration.

An important outcome that may not have been identified in identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.

Bruce Steinwald: Comments?

Male: I know that these are our own criteria that we wrote ourselves as a committee. But in my own mind, I wonder that the hurdle of accuracy that one might use for quality improvement might be quite different than the hurdle one would use for public reporting.

An example I might use is some of the arc measurements where they were validated at the 50% accuracy rate using some of the coded data. And one view of that is, hey, 50% is good enough for the public because it's better than knowing nothing. But another view would be that it's equally probably wrong as right. And most of the time equally wrong is right wouldn't be the hurdle that we would want to use in any sort of scientific way.

Joe Stephanski: That's always been a problem with some of the discussions I've had in the years gone by and few after that. Business decisions are basically made on a 60 - like a two thirds likelihood of being correct. And that's not a clinical standard.

So you run into this dilemma of what works in the clinical world versus what works in a decision making world in business.

Bruce Steinwald: Well, and this is the intersection of those two worlds. And 50/50 would not usually cut it. For me, although, you might be able to use that even in quality improvement. Because if it's not - people are going to be publically embarrassed. They might be willing to act on data that isn't so good but they'll accept it generally because they're all trying to move in the right direction.

At the moment at which they're going to be publically embarrassed it becomes a different ballgame.

Sue Knudson: What is drawing your assumption that this is, like, only 50% accurate?

Bruce Steinwald: I'm using some of the other measurements as an example. Although I would say that I think the risk adjusting methodology and the discrimination between 5% better and 5% worse which looks like the spread in the data means that there's at least a 50% chance that those rankings are not completely accurate.

They're probably accurate enough. They're probably directionally accurate but they're probably not there's - at least in my mind - a 50% chance that they're not exactly accurate.

But I'm willing to over - I'm not making that assertion specifically about this one. I'm making that assertion sort of more generally. And I get the sense we're spending perhaps more time on this one then we might because it's our first, you know, big analysis of one of these. And therefore we're going back and checking against our own methodology or our own criteria a little bit.

Joe Stephanski: That's certainly - I hope that's true. Because I don't think we can afford to spend so much time on everything. Although we only do have one other non-condition specific measure.

I don't interrupt the use of ability criterion to mean that the measure would have to be usable for every purpose that one could conceive of. In other words, if you were - if we, collectively, uncomfortable with using the measure for public reporting. That itself would mean the measure was not usable.

Bruce Steinwald: It would not mean it was not usable.

Joe Stephanski: To me, that's my assertion.

Male: I'm completely comfortable with that but as we've constructed 3B, it lumps all the uses together.

Where I do think there's some discrimination about the uses and I just don't know that we've set up our own criteria to make that discrimination. If we wanted to, and I'm posing the question of if as a group we even want to.

Doris Peter: And I, this is Doris from Consumers, and on previous committees I've been on there was an explicit direction that measures have to be sufficient for public reporting and quality improvement.

So I wasn't sure, if that, that doesn't sound like it agrees with what you were just saying.

Male: Right.

Male: I think that's what we had said at - that's why 3B reads the way it does.

Male: Right, right.

Male: But in relationship to this one for example or perhaps others similar to this, there could be a distinction. And what you're saying is, I'm like the chairman. If it isn't suitable for all purposes, we should vote no.

Doris Peter: I mean that's just how the criteria is set up, so that was my impression of how we were supposed to approach the measures.

Male: It is, it is...

Barbara Rudolph: This is Barb. From the (C-sect) perspective measures should be appropriate to public reporting and quality improvement with the caveat that sometimes a measure that's publically

reported may not have explicit sort of uses in quality improvement. That sometimes it's to notify providers that there's an issue here and the providers need to figure out a way to improve the quality.

Heidi Bossley: This is Heidi, just to interject for a second because I know you're talking criteria. You really do need to determine whether or not it's suitable for accountability. That's truly what's on the table for you to consider today.

Joe Stephanski: I'm sorry who was that that just spoke?

Heidi Bossley: It's Heidi, ((inaudible)) staff.

Joe Stephanski: Great.

Sue Knudson: And I think the other thing that's a consideration is it sounds like from this measure's actually being reported out. If not to the public at large, but to members of Health Partners and has been for sometime. So, and that's consideration too.

Bruce Steinwald: It's already being in this matter.

Sue Knudson: Already being used.

Bruce Steinwald: Yes.

Heidi Bossley: Right, this is Heidi. I would think at the three year review we would look to see how much better that reporting has progressed beyond just within that framework of that group. But this is actually probably more use than what we typically see. So.

Bruce Steinwald: This is Bruce again. Remind me again what - it's purposeful that we have the word quality in there twice - quality improvement twice. We decided not to change that to resource use improvement. Is that true?

I admit to always being a little bit thrown by trying fit resource use into an existing framework for quality. But I just want someone to verify that, yes, we intended to say quality improvement here.

Ashlie Wilbon: Hi Bruce, this is Ashlie. Yes, we tried not to wordsmith too much. But essentially what we landed on - I think kind of between steering committee discussion and QF internal discussion is that quality improvement, particularly in the context of resource use is a more broader umbrella. So we're just looking for general improvement and just general improvement and quality of care.

So, it's not just clinical quality.

Female: Right.

Ashlie Wilbon: So I hope that helps.

Bruce Steinwald: Okay.

Ashlie Wilbon: Or was that too broad? But it's a broad - it's a broader definition.

Bruce Steinwald: Right, thanks. Well now I'm not sure if we're at an impasse or not. I mean to mine mind I would like to leave up to the user what particular uses the measure could be put to for their purposes.

And, you know, it - this risk reward kind of calculus would need to be gone through every time and whether they felt it was too risky for public reporting or not, I think would be up to the user. I don't know why that would be up to the steering committee. But others can disagree.

Sally Turbyville: And Bruce, this is Sally, just to add to that. What we're asking the steering committee is whether it would be suitable as specified for public reporting. But you're absolutely right that a user of this measure would be endorsed would want to weigh how they would actually roll out the use of the measure.

And that would not be written in detail. It's just a matter of - is the measure suitable through interpretability through it's construction. Ability to produce a score that would be meaningful for public reporting. Whether or not it's actually publically reported, as Heidi mentioned, we would re-evaluate how much it progressed in that - in the maintenance in three years component of this project.

Male: So, if I could just clarify. Is the threshold 60% accuracy? If we could put such a number on it. Was it at least suggested sort of as we say sufficient for business purposes?

Ashlie Wilbon: So, this is Ashlie, we don't necessarily put a number to it or a threshold to it. I think that's kind of what we're looking for the steering committee to decided based on what you've learned about the measures so far. And how it specified, like Sally said, are you comfortable with how the results would come out for it to be used for public reporting?

Male: Well, but that's - that's exactly the question I'm asking. But I appreciate there's no specific number that gets put on it. But if generally, this is likely to be 50 or 60% accurate, is that the threshold which we're determining that something ought to be publically reported?

Or do we believe that there should be some higher standard?

Bruce Steinwald: Like others, I'm unsure how to respond to that. This is Bruce again. Once again, I would think that should be up to the user and depend on the uses to which they want to put the measure.

To say 50 or 60% accurate to me doesn't give quite enough information. Let's say if they want to identify groups that are above a threshold. They can set the threshold as high as they want to. Right?

So I don't - so that the, you know, the chance of the provider being unfairly singled out would be very, very low. But that would be their decision not the measure developer or ours.

Well, the silence once again is deafening.

I just will restate my view. I think this is up to the user. Not so much up to us if we think that the measure is suitable for public reporting and other uses.

Paul Barnett: This is Paul. I'm a little bit confused by the discussion because I don't think we have that number in. We have this other criteria that we worked out. So I'm not quite sure why we're at this point.

We don't have the number and we already didn't anticipate having one that we would use to make a decision, so.

Male: Right, so, come to a conclusion then. Then that says to you that the issue that we're discussing is really germane to whether or not we would recommend ((inaudible)).

Paul Barnett: Yes, I think that's right. I think we just need to go through the criteria that we, you know, use the criteria. And then apply them and come up with a summary.

Male: Okay.

Paul Barnett: Whether yes or no.

Male: Okay.

David Redfearn: This is David Redfearn. I mean I didn't look at this from the point of view of, you know, how accurate the method would be and how much error would be built into it.

I looked at it from the point of view of complexity in terms of public reporting. And my reaction, this is enormously complicated. And when I talk about public reporting I think of the quote public.

And I think it would be very difficult for them to understand and appreciate these scores when they come out. I think professionals, physicians and analysts and people like us could figure it out. But I sure wonder whether the public could figure this stuff out. And in that sense, I think it is probably not appropriate for public reporting.

I think my response was that it had low - that was low in terms of public reporting. But I looked at it from the point of view of how complicated is it, how easy is it to understand the methodology.

David Penson You know, let me add to that though. I mean, David, I think this is David Penson, I think you're right to some degree. But I think we're going to run into - I mean then we're not going to be able to approve any of these. Because the Ingenix measure - in my - and I know you can't compare them that's not fair.

But they're all very difficult and I don't think the general public grasps these. It's hard enough for me as a clinician that does health services research to grasp these things. So I don't know what the ((inaudible)) for that.

Male: I agree. I mean my react - I know the Ingenix stuff inside out and it still strikes me as enormously complicated.

David Penson At least I felt like I could get my arms around this one a lot easier than I could around the Ingenix one. And, again, it's not - you can't compare them. I felt like I could make heads and tails of this. But I think your point's well taken. Your average Joe on the street and I don't mean to sound condensing. This isn't going to make sense to them.

But if that's our sort of requirement then I don't think we're going to be able to get any of these approved.

Female: And this is - oh, go ahead, Barb.

Barbara Rudolph: This is Barbara Rudolph. Well, you know, I read Consumer Reports and if they are talking about a vehicle. I may not know how the engine exactly works. But if it's a legitimate source and they give me some symbols I can understand that. And I think it's - we can not expect that the consumer's going to understand everything about the modeling or the risk adjustment.

But they will understand that we accounted for the fact that some physicians see patients who might be sicker or have more conditions. You know, I think they don't - the consumer doesn't need to know the exact workings of this in order to have a benefit from it.

And I think I'm happy with seeing some dollar symbols and other kind of things like that. And trusting that my source is accurate.

Joe Stephanski: Well, I accept as well frankly. At least for this measure.

Doris Peter: Hi this Doris from Consumer Reports, if I can chime in on what we've been doing. Actually a few people are speaking in it and the evidence that we have so far complete supports your view that consumers don't understand this at all and we've been struggling immensely with trying to present this information. Because, as you all know, that most consumers will think that, you know, more cost is better.

That they would say, oh, this hospital uses more resources so I want to go there. And we battled with that for several years now. We've been trying to present the Dartmouth Atlas Data and other data.

But to Barbara's I think it's up - it's actually up to us, the Consumer Reports to take that information and translate it in a way that consumers can understand. That's what we'll continue to do. But we would like to use measures that NQF approved or at other endorsed and that are well vetted. So that we know that we trust the data that we're using. And then we can translate it in a way that consumers can understand.

So I know that's sort of a half-half answer but, you know, we need this information to be able to make it understandable to consumers.

Bruce Steinwald: Right the other factor is it's more understandable to the providers. And it's more meaningful to the providers when it is publically reported. So we're not just relying on consumers to make decisions based on the measures. It will spill over to the providers too.

Anything more? Go ahead.

Ashlie Wilbon: So, I was just going to say it sounded like everyone - I think we kind of wrapped (3C) in that discussion as well. Kind of talking about how well or how transparent the measure is and how understandable is - the results would be.

But 3D is we're actually - it's going to be not applicable for this point in the project. Where that harmonization is something we'll address later as need be so we can split that one for now and move on to feasibility if you're ready.

Male: Okay.

Ethan Halm: This is Ethan. I have a question. I mean in some of these tensions where everyone's struggling with about, you know, sort of the how - the interpretation of this. And what would happen if, you know, a government regulator would crank this through and make it available to the whole public.

Are there any lessons to be learned from all the quality measures groups as far as how they've dealt with the idea that none of the things are perfect. They're all complicate. You know, interpretation's in the eye of the beholder. And do we need to worry about certain interpreters or certain interpretations.

Or are these, you know, have to be deemed, you know, good enough for prime time. Or the best that we can do, you know, but in a way that's justifiable and documented. And then sort of move on.

Because, obviously, people worry about these a lot more, I think, than the quality measures.

Male: It really depends on the manager and it depends on the circumstances. But there will - there's always been a tension between the consumers, purchasers, and the clinical community. So we have to balance those tensions.

Male: So, like, for example with the quality measures are there some where people say, well, these are fine if people want to use them internally for quality improvement. But I wouldn't want Medicare or Medicaid or the Cross using these and publishing them or sending it to members.

Male: Happens all the time.

Male: Yes, I mean that's accountability versus a QI measure.

Bruce Steinwald: Let's keep in mind, we've said from the beginning that this is an interim set of measures on our way to linking quality and cost and looking for efficiency and effectiveness in the delivery of care.

We're not going to be there at the end of this. And yes the concern is real. Consumers are going to say more is better. But we're not going to get to the endpoint that we hope to measuring efficiency and effectiveness as front.

Male: Well, I guess the question is down to - when you look at some of stuff like Ingenix did in some of the local areas. The question is the confidence intervals on the data. And how fine a line do slice on people?

So do we have to put in somewhere or do we put in somewhere that it's hard to make distinctions between providers within a certain band of difference in score or difference in performance. And is that part of the equation that we may need to have as an explanation about the use of the measures.

Male: Mike, at some point we're going to have to raise that issue. At some point we're going to have to deal with the whole issue of hierarchal modeling whether we should be biasing low volume providers toward the mean which is a complicated issue.

But the - but I think we come to that when we've evaluated the measure itself. And we're, you know, to some extent we're the issues ((inaudible)) is raising in the call right now are about whether to even go down this route which I thought was a decision we'd already made.

Knowing all the limitations that we're currently talking about. Consumer uncertainty, who the user is, the fact that's an incomplete measure because it only measures resource and not ((inaudible)).

Bruce Steinwald: This Bruce, that's right. I think that was our discussion early on at our one face-to-face meeting. We kind of came to that conclusion. The only other thing I would say though is that we don't - if we endorse a measure I don't think we want to undermine the value of the endorsement by pointing out its limitations.

But nevertheless, I think we have to look at the half fullness of the glass of what we're trying to accomplish here and then forge ahead.

Anything, where are we Sally? I guess we're still at 4A or we're at the top of feasibility, right?

Sally Turbyville: So, feasibility really looks at in 4A of the data are routinely generated. As you know these measures are based on commercial claims data.

Male: Right.

Sally Turbyville: Whether the required data elements are available in electronic, they are electronics, the first two typically have gone quickly for the resource ((inaudible)) measures. And then, you know, kind of the susceptibility to inaccuracies and errors I think you all have talked about that quite a bit. And the unintended consequences and the importance of monitoring those.

And then the data collection strategy, again, is a rather forward one in that it's using the claims data however what is the twister for 4D is the fees associated with the ACG approach. Because it does contain a proprietary specification in there so kind of tricky ones for this group is 4C.

And then 4D is thinking about if the fees associated with ACGs constitute such a large barrier that it makes the feasibility lower or not.

Bruce Steinwald: Any comments?

Female: ((inaudible)) and one of my thoughts that have been rumbling around my head through all this on the feasibility of applying these measures outside of the Health Partners arena is that the commercial data set are not generic. I mean we have all the coding and stuff like ((inaudible)).

But how our data warehouses are set up for different types of products. What kind of data's required for payment particularly large self-insured accounts. All ((inaudible)) that this is somehow a standard generic thing out there. It is for my working and trying to understand data within my own organization is a bit of a stretch.

Now I don't want to, I mean for me also the biggest question of this whole enterprise is how are these measures when applied as they're written effective in getting us to better efficiency of care? And so if my concern are legitimate from a technical standpoint but not significant from an impact standpoint. I can be quiet about that.

I just want people to know as a commercial payer insider that this is not a generic commodity this data.

Bill Golden: This is Bill Golden. And we've been working with this with Blue Cross in Arkansas as well as Arkansas Medicaid. And there are increasing protocols and methods for creating all payer databases where there is a way of basically taking separate databases and homogenizing them based on certain claims file and definitions.

So in fact we're going through that right now. Where we're going to sharing and merging some data. And different third-party administrators can cleanse and fix up their data sets to match up with each other, certain codes and certain categories of codes.

So that's increasingly common kind of activity that I would hope would not become a barrier for this kind program as long as you can do the crosswalks.

Female: Right, this a lot of variation, regulatory variation from state to state I think about how much leverage that we can apply to the self-insured account in terms of curing data that truly belongs to them. You know and so we end up maybe measuring smaller percentages of individuals' practices.

But I know that there are things that are happening in various markets and maybe we'll get a more common approach to this. We're not there now so would be my point.

Bill Rich: I think these are all important points, this is Bill Rich again. And to follow up on Dr. Golden's comment. This is a very unique patient population. They're mostly large groups seems to work for the large groups. And I'm not sure it is applicable to self-insured at all. So I think our description of this should be that this was described and how it's been used and how it was presented in the patient population.

I don't think we can generically bless this. Or, you know, an employer with an area with a lot of solo family docs and things like that.

Does that make sense Bill?

Bill Rich: Yes.

Bruce Steinwald: Any more comments or questions on feasibility?

Paul Barnett: I guess the - in some of the summary of what some of the earlier stuff was said, this is

Paul. The concept that has only been done in one part of the country with one plans data would be a little bit of a concern because, you know, of what people are articulating. That it - we don't know how well it'd work with other data.

Although, on the other hand it seem fairly straight forward to work from the claims data to translate to these things. It seems eminently possible. But that is a little bit of a limitation and it hasn't been widely applied yet.

Bruce Steinwald: If we were to say it's feasible for use in large groups do we need to define what we mean by a large group?

Female: Bruce, was that a question?

Bruce Steinwald: That was - I was throwing out a question for everyone, staff included.

Tom Rosenthal: This is Rosenthal, I think it's fundamental question. If the word, no it only makes only sense in groups that have more than six primary care physicians in a group, then you would have to incorporate that and say it's only applicable in those settings.

Barbara Rudolph: This is Barb, I don't think we can hide that. I think, you know, it came in for endorsement the way it is and obviously anyone looking at this measure and doing any sort of due diligence on it would apply to them or not.

And there's certainly lots of large groups out there. I live in Wisconsin and we're all large practice groups. California has, of course, Kaiser a very large practice group. I mean there's lots of other places that this can be used. I don't think we should hide anything that we don't have evidence for.

Carlos Alzola: This is Carlos. I think they're submitting the initials for providers with more than 600 members.

Male: Okay, that was in the specifications? Was it?

Carlos Alzola: It was in the, I think, in the ((inaudible)) section. I think.

Male: But I thought it said it was only applicable though with groups because of the attribution question. And therefore I appreciate we can't add some restrictions that aren't there. But it does say that it only discriminates between groups. But it doesn't say how large the physician groups have to be.

Carlos Alzola: That's correct.

Male: But it seems germane to the attribution problem.

Male: Well, then where does the 600 members come in then? I'm missing something.

Male: Like that was groups that have fewer than 600 members.

Male: Okay, all right.

Bruce Steinwald: Any thing further?

Ashlie Wilbon: So, if everyone is okay and they feel there's no more questions in order for them to - for you guys to complete the evaluations that we'll send out later. I will go ahead and transition to the Ingenic measure when you're ready.

Male: What should we do to get ready? Should we stand up and stretch?

Female: Yes, you can do some jumping jacks at your desk and I'll just ask if Cheri's there, Cheri Zielinski from Ingenic? Operator, can you see if Cheri is on the participant line, by chance.

Male: What's Cheri's last name?

Female: Zielinski.

Male: Yes, Cheri's line's open.

Female: Her line's open?

Cheri Zielinkski: Hello?

Female: Hi is that Cheri?

Cheri Zielinski: Yes, hi.

Female: Hi, Cheri. So we're going - I know we're a little behind schedule but we did want to get started on the Ingenix measure. So I just wanted to make sure you were there. And to see if you could provide just a brief overview for the committee before they start discussion on the importance criterion.

Cheri Zielinski: Sure, so the measure is a population based measure. It's not - it's not condition specific. It's used to measure populations. And , you know, it looks at utilization and observe two ratios of those populations.

I think that's about it.

Female: Okay, thanks Cheri.

Cheri Zielinski: Sure.

Bruce Steinwald: Who would like to make some statement or ask a question of relating importance?

I'm finding it hard to see what it would be any less important than any other, you know, population based measure. But I could easily be missing something.

Male: Well that's high importance still.

Bruce Steinwald: Yes. Anyone want to argue for low or even moderate importance?

Paul Barnett: So, I'm - this is Paul Barnett. The question was, so the - is - and it seemed like the other submission was also confused on this. Not that the - not that the - I thought the meeting of this criteria was not that the question of health care costs is important. But rather that this measure itself has had an important effect.

Male: Has had?

Paul Barnett: Yes, I think so.

Ashlie Wilbon: So, this is Ashlie. Just to clarify, the importance criteria, sub-criterion are really focused on the measure focus. So the topic of the measure, in this case it is the topic of the measure is to measure cost and resource use. Measure across a non-condition specific population. So, that would be what we're focusing on in terms of whether or not that is important. If that makes sense.

Female: So it's first starting with is it important to measure resource use in a population based approach that this measure looks at. Is the area in which they're measuring which is a per capita measure, is it a high impact area or evidence that there's - I think often what we talked about in the steering committee was variation being an important part of why we want to measure.

Is it an area where there's a belief that there's opportunity for improvement. So is there a way to reduce that variation. Not so much how the measure would perform within the focus area. Then 1C we start talking about the measure intent. So are they describing in a very coherent way what the intent of the measure is.

And then the last one is again talking about within the measure. Are they describing the intent of the resources that they are capturing, is it described well. So it really is about the area in which they're choosing to measure. And then when we go into specifications that's when we get into more detail about the measure itself.

So I think the big question at the high level, is it important to measure resource use at a per capita approach? And is there evidence that resource use variation has an opportunity to improve etcetera. And they would garner that evidence through literature review or because it has measures for a while they provide some information of the variation that they've seen, etcetera.

Male: Although this particular one is looking at not per capita but per episode approach. Right?

Female: No, it's a per capita in base measure.

Bruce Steinwald: I guess you don't have that on the screen but I think it is a per capita measure, isn't it?

Female: It is.

Female: It is a per capita measure. That is correct.

Male: Well, okay.

Female: Any questions about is this an important area to measure? Any weaknesses? You know why you might think it's only a moderately important area to measure per capita resource use? As Bruce was asking.

Bruce Steinwald: Let's move on.

Female: We move on, okay.

Female: Well, we weren't going to dive into the scientific acceptability of the Ingenix measure today because we had an assumption that the Health Partners measure would take some time and we

didn't want to break up the scientific acceptability too much. So unless there are any other questions about the importance of the Ingenix measure or if you have questions of Cheri from Ingenix that you think might help think about the scientific acceptability which we will assess in our future call.

Or want to go back to the Health Partners measures. We can convene - adjourn this meeting a little bit early.

Bruce Steinwald: This is Bruce. I would personally be grateful to be able to do that.

Male: God bless you Mr. Chairman.

Lisa Grabert: Sally, this is Lisa Grabert. Can I just ask a clarifying question of Cheri?

Sally Turbyville: Please.

Lisa Grabert: Cheri it was hard for me to tell from the documentation if the per capita measure was calculated by accumulating of the individual episodes that were attributed to one patient or not.

Cheri Zielinski: It is - the answer is yes. It's accumulating all resource utilization for a member. So it's member based and not episode based.

Lisa Grabert: Okay does the documentation refer to what would happen to claims that were in a ungroupable episode. Any individual services that were not able to be grouped into one of your episodes. What happens to those in the overall per capita measure?

Cheri Zielinski: I'd have to take a look and I'm not sure if that's a specification. I think it's in the specifications. Ungroupable records are not considered as - they're exclusion criteria for the

measure. So for these measures I believe we're looking at only complete episodes. And those record grouped episodes.

Lisa Grabert: And do you know on average how claims, the percentage of claims that actually group for ETGs.

Cheri Zielinski: I do not up off the top of my head. No, I , we...

Female: Well take that as a note stop for a follow up Lisa.

Lisa Grabert: Thank you Sally.

Sally Turbyville: Any other questions about the Ingenix measure or it's importance.

Female: Maybe besides - it's a follow up question on how many of the claims do or do not group. You know, so how much would be excluded from the analysis? It might be helpful to know if we have any kind analysis here or something about the ones that are not accruing. Are they, you know, significant clients or not significant clients.

Sally Turbyville: Okay.

Female: Thanks.

Sally Turbyville: Of course.

Bruce Steinwald: We reconvene or by telephone on June 22 is that right?

Sally Turbyville: Yes, just a couple of things before we adjourn. We do need to do a public comment period. Briefly, Robbie if you're there if you could just queue the audience for anyone who might have questions for public comment?

Operator: Absolutely, for those who do have public comment please press star 1 on your telephone key pad at this time. If using a speakerphone please make sure your mute function is turned off to allow your signal to reach our equipment. Again, star 1.

And we have nobody in queue at this time.

Sally Turbyville: Okay, great. Thank you. So before we adjourn, I just wanted to go over a couple of upcoming activities and dates. We will be sending out a link today for everyone while it's fresh on your mind to do your final evaluations of the Health Partners measure. And submit your vote for whether or not you think it should be recommended for endorsement.

So that will be sent to you today with any instructions based on questions that were asked today to make sure it's clear for everyone so we can get those submitted hopefully by the end of the week.

We will actually - we had a call on the calendar from a long time ago that we're actually going to go ahead and use to wrap up the discussion, hopefully, of the Ingenix measure before you come to town at the end of June. That is June 22 from 12 to 2. We'll be sending out dial in information and an agenda for that in the coming week or so.

We will also be, we're getting the last bit of information from the developers for the CV Diabetes measures that were reviewed by the CV Diabetes tap a couple of weeks ago. Those are the measures that will be the focus of the in person meeting when you come at the end of the month. So we'll be sending you those measures along with the tap, you know, evaluation of those

measures and comments and ratings for your review. So that you can begin to prepare for the meeting at the end of month. We'll be sending those this week as well.

You should have also received an email from our meetings department with instructions and information on how to arrange your travel for the in person meeting. If you have not received that or if you have any questions about that please feel free to email me or (Camille) who you should've received the email from specifically about making those arrangements.

So I think that's it. And if anyone has any questions. Okay, if everyone's okay, then we'll go ahead and adjourn early today, 15 minute early.

END