

## **NATIONAL QUALITY FORUM**

**Moderator: Ashlie Wilbon**  
**July 14, 2011**  
**2:00 pm CT**

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

Ashlie Wilbon: Good afternoon everyone. Welcome to the CV/Diabetes Resources TAP Conference Call.

I want to thank everyone for joining us this afternoon. And I'm going to just go ahead and jump right in.

Can I just get - do a quick roll call from the TAP just to see who we have on the call? So when we get into discussions, we kind of have an idea who's out there.

Jeptha and Jaime, you're there?

Jeptha Curtis: I'm here. Yes.

Ashlie Wilbon: Okay.

James Rosenzweig: Yes.

Ashlie Wilbon: Mary Ann Clark?

Mary Ann Clark: Yes, here.

Ashlie Wilbon: Okay, great. Michael O'Toole?

Katherine Reeder?

Dr. Kay Reeder: Yes.

Ashlie Wilbon: Brenda Marie Parker?

William Weintraub?

William Weintraub: Here.

Ashlie Wilbon: Okay. Thomas Marwick?

And David Palestrant? Okay. And David was going to join us.

Okay. So thanks everyone for joining us. Today, I wanted to provide a general project update -- we had some activities going on since we last had a call -- just to kind of give you an idea of how the measures that you've reviewed went through the steering committee meeting that we had a couple of weeks ago, just provide a quick evaluation process refresher before we jump into evaluating the measures.

And I know we sent out an agenda I think earlier this week or late last week. We've done kind of on-the-fly adjustment based on a new development that we'll talk about. But today we're going to focus on getting through the evaluation of 1591 and 1594, both from Ingenix. We do have the

developers on the call to ask questions. We also have Carlos Alzola on the call who's our statistical consultant. So he can also offer some insights along the way.

We wanted to try something different today on the call because for those of you that are at your computers, we wanted to have you open the SurveyMonkey during the call, and as we're discussing the measures, have you actually rate it in real time and submit it to us when the call is over. We are having a steering committee call tomorrow, and we wanted to have the results of your discussion today to be able to present to them tomorrow.

So to the best of your ability, if you have access to a computer, we're going to ask that you try to bring that link up on your computer during the discussion so we can kind of facilitate a quick turnaround there.

Can I just get an idea of who is in front of - is everyone from the TAP near a computer right now or are some people just dialing on the phone?

Mary Ann Clark: I'm at a computer. This is Mary Ann.

Male: I'm at a computer.

William Weintraub: This is Weintraub, near a computer.

Ashlie Wilbon: Okay, great. Let us know if you have any difficulty. We'll prompt you kind of throughout the call, during the discussion if we want to say kind of we're at a point, or if you feel comfortable during the discussion at any point to submit your ratings, that's fine too. And that will be greatly appreciated.

So I just wanted to give a quick project update. In addition to this TAP, we have three other TAPs. The Cancer TAP met on June 28. They had four ABMS measures and a one-day meeting. And then they got through all four measures. Things went really well, and Dr. David Penson chaired that TAP.

And they had - they're quite different from the CV/Diabetes measures. We had two breast cancer measures and two colon cancer measures. And, you know, some interesting things were identified. But they were very different somewhere at the regional attributed, or the level of announces is at the regional level, not the physician level. So, definitely some differences in there.

The Bone Joint TAP met last week. They also had four measures -- two from Ingenix and two from ABMS. One was on - the ABMS measures were on low back pain and radiculopathy. And we had a knee - hip and knee replacement and a hip and pelvic fracture measure from Ingenix.

The Pulmonary TAP will meet July 19th. They will have about six measures - five measures to go through from Ingenix and NCQA. And we're looking forward to that meeting coming in town as well, so hopefully wrapping up these meetings very soon.

The steering committee met on June 29th and 30th. And, Jephtha and Jaime, feel free to jump in at any time here with my recap here. But they had a two-day meeting and they ended up recommending four measures, two of which were non-condition specific measures both from health partners. One is a total resource use, population-based measure and the other one is a total cost of care population-based measure -- so one uses standardized pricing and the other uses actual cost. And then they also recommended RRU diabetes measure and the RRU cardiovascular measure from NCQA.

I will note that the discussions on these measures were very interesting and the votes weren't always - there's definitely a lot of support, a lot of I guess less division among the committee on the NCQA measures. The total resource use measures and the total cost of care measure definitely created a lot more discussion and division among the committee. There was a very narrow margin of yes/no votes for a scientific acceptability and for recommending the measures. So those all will go out for comments and we'll see what happens. But they will continue to move through the process as recommended.

A relatively new development in the project is that ABMS has decided to withdraw the measures - all their measures from the project at this time to do more testing.

Male: Oh.

Ashlie Wilbon: And so what we are going to do at this point is use the feedback that we've already gotten from you, obviously compile that and send it back to the development group so that they can, you know, use it in their process to continue to refine the measures and continue testing.

So for this call today, if we are able to get through the Ingenix measures, we will - this will be our last call. And what we'll do is for the primary reviewers of the last two ABMS measures for 1574 and 1575, we'll e-mail you to get any notes you had and input. We'll e-mail the whole TAP. But particularly for those who are the primary reviewers, we'd really like to get your input on your review of those measures so that we can compile it and feed it back to the developers for input as they continue through the process.

I think what we found through the process is that everyone really liked the measures, but they didn't feel like they were quite ready.

So any questions or anything to add, Jephtha or Jaime, about the review of the measures at the steering committee meeting?

Jephtha Curtis: No.

Cheri Zielinski: Oh sorry. This is...

Jephtha Curtis: No.

Cheri Zielinski: This is Cheri from Ingenix.

Ashlie Wilbon: Yes.

Cheri Zielinski: I just wanted to make sure I was clear that the Ingenix measures, they were not - there was no final vote on the steering committee is that correct?

Ashlie Wilbon: Yes, that's correct. Sorry, Cheri. So tomorrow, we have a steering committee call to discuss the measures we're discussing today, the 1591, the 1594. And we'll wrap up discussion on the two measures that'll be on the agenda for the steering committee meeting from Ingenix, which is a non-condition specific measure, as well as a diabetes measure.

Jephtha Curtis: Actually, this is Jephtha.

(Crosstalk)

Jephtha Curtis: I just wanted to follow up on what you said that the steering committee I think really appreciated the deep-dives that we've taken on all these measures. And that although it sometimes felt like we were, at least to me, shooting in the dark a little bit, I think we really hit on

the key issues for all the measures that we evaluated. So I just want to thank everyone for their efforts. And I know that they were very rewarding at the steering committee level.

Ashlie Wilbon: Yes, definitely. Thank you, Jephtha.

So that said, I think let's go ahead and move on. For today, we have about 45 minutes per measure. And when we just - before we jump back in, for 1591, I know initially on the agenda we had said we were going to only discuss usability and feasibility. But when we went back to the transcript and to the recording and our notes from the call, we realized that the discussion of that measure was really - it was towards like the last few minutes, maybe the last 20 minutes of the call and we didn't probably do as thorough review as we probably should have.

So we'd actually like to do a quick recap of the discussion that we had on 1591 on the last call. And I can bring that up on the screen. And then we'll just start with scientific acceptability again. Kay will help lead us through that. And then we'll move on through the usability and feasibility for that measure and then move onto 1594.

So we've got about 45 minutes for each measure. And again, we're going to try specifically for the 1594 which will be our first time discussing that measure, we'll have the co-chairs lead us through a brief important discussion. And then for scientific acceptability, we'll have the lead discussant lead us through and highlight any issues.

We also have Carlos on the phone who can give a brief overview of his assessment of the scientific acceptability of the measure. And then we'll move onto the usability and feasibility.

So again, we're going to do a systematic review of each of the criteria. We'll go sequentially and highlighting any areas that need to be discussed.

And I just wanted to do another quick review of the evaluation of the scientific acceptability particularly around the reliability and validity, and what exactly delineates a high, a moderate and a low score.

So for a high reliability and validity rating, what you're saying is that all the measure specifications are - they're unambiguous. They're completely consistent with the intent of the measure and consistent and unambiguous in a way they've identified how cost and resources should be measured and how the score should be computed. And that they have evidence of both - reliability of both the data elements and the measure score. So again, it's a very high threshold there for getting a high score.

For the moderate, for reliability and validity, the specifications are unambiguous. They reflect - again, for validity, they reflect the intent and - cited in the importance to measure, and that they've done - they've demonstrated reliability of either the data elements or the measure score of the measure. And then for validity, that they have done a systematic assessment of face validity or that they've shown a validity within acceptable norms of the measure score or the data element and identified any threats to validity as well.

So for a low reliability or validity, you're saying that the measure specifications are ambiguous, that there is potential for confusion and that for validity, that the specifications don't reflect the evidence or the intent of the measures cited in the importance to measure report. So again, there was low - you're saying that there's some ambiguity within the way that the measure specified and the level of reliability and validity that's demonstrated.

And then insufficient, obviously, if you don't feel like that you have the information you need to determine whether or not it's a high, moderate, low, again you would use insufficient.



So just a real quick recap, I think I already talked about all these items. But we will be trying to get input on the remaining ABMS measures via e-mail for 1574 and 1575 from the entire TAP. So we can forward that onto ABMS. And then again, we'll be trying to do real-time results of your voting today so that we can pass that onto the steering committee for their discussion of 91 and 94 tomorrow. And then if all goes well, today will be our last CV/Diabetes TAP call.

So again, thanks to everyone for your efforts. I know it's been a long road particularly for this group. And we appreciate your perseverance.

So on that note, if I could hand it over to Jephtha and/or Jaime to kind of get the group started, that would be great.

Or, Kay, if you want to - or why don't we do this, I'm sorry. Why don't we have Ingenix do a brief introduction for the measure? And then we'll go to Carlos for a brief overview of his assessment. And then we'll start with the TAP discussion.

Male: Do you want me to reintroduce congestive heart failure?

Ashlie Wilbon: Yes, that will be great. Thanks, Tom.

Thomas Marwick: Okay. The congestive heart failure measure from Ingenix, that is a measure that gathers claims from the administrative claims stream into an episode of congestive heart failure. And because congestive heart failure chronic disease is a viable episode, it's a yearlong episodes of congestive heart failure, but then...

Ashlie Wilbon: Okay. Is that it, Tom?

Hello? Hello?

Hello?

Male: Hello?

Ashlie Wilbon: Hi, yes. Can you guys hear me?

Male: Yes.

Ashlie Wilbon: Oh okay. Is Carlos there?

Carlos Alzola: Yes.

Ashlie Wilbon: Okay, Carlos, do you mind giving just a brief overview of your assessment of 1591?

Carlos Alzola: Sure.

Ashlie Wilbon: I think you did it before. But it'd be great to just have a recap.

Carlos Alzola: Okay.

Ashlie Wilbon: Thank you. So we focused first on - they focused first on the summary assessment. They

- I felt that clinical and construction logic of the measure was described in sufficient detail. And the measure users should be able to implement it to using the description provided.

The most - one of most important parts was to look at their reliability and validity. And the reliability was established. And they also established - in my view, they also established a face validity.

The methodology was applied. This can be used for a group of practice, individual clinician, other clinician teams, facility, health plan, at the level of integrated validity system, county or city, state, regional or national level.

In terms of things that I would like to see more detail is in the statistical models, and that there's no details in terms of the ((inaudible)) of the models or - and especially the calibration.

So I can go into a little more detail into their reliability or shall we - or is there something else you'd like to me say?

Ashlie Wilbon: Yes. Why don't we just have the TAP - Kay, are you there?

Dr. Kay Reeder: Yes.

Ashlie Wilbon: Could you maybe recap the discussion that we had on the last call on this measure? Do you want to just kind of - I have the notes, I think, on the screen that I sent you. But maybe you want to just kind of recap for the group what the discussion was a little bit.

Dr. Kay Reeder: I don't have that screen up and...

Ashlie Wilbon: Oh okay. I can do that.

Dr. Kay Reeder: Okay. Thank you.

Ashlie Wilbon: So just to recap for the discussion that we had on 1591 on the last call, the - you had the important discussion. The TAP was in agreement that the important criteria was met, CHF is a

high-impact condition and a high cost area. And this measure basically covers an episode treatment group across the 12-month period, with the claims and dollars associated with it.

With the scientific acceptability discussion, there was a discussion obviously about some of the specifications. There seemed to be an agreement that the specifications were precise, that there have been extensive internal benchmarking and comparison that were conducted, although the TAP would have liked to see more external comparisons; that the data of - for testing across - that was submitted was from across nine health care organizations, all from the large commercial insurers and were geographically spread.

That - I think there was actually if I recall, a big discussion on systolic versus diastolic episodes and codes, and if the population identified by this measure included a mix of both, and how the measure was being defined and how the two, systolic and diastolic heart failure, were being grouped in the claim.

Female: Ashlie?

Ashlie Wilbon: Yes?

Dr. Kay Reeder: I think the way I recall that discussion, and someone can correct me if I'm wrong, was that in the beginning there was some confusion about the terminology "congestive heart failure," which is also NQF terminology. And that some - it was brought up that some heart failure is not congestive and that there needed to be more clarification there. And then the discussion went on that, indeed, this is mainly a systolic and then a mix systolic/diastolic heart failure measure and that Ingenix also has a diastolic heart failure measure.

Is that correct?

Ashlie Wilbon: Was that to NQF staff or to Ingenix?

Cheri Zielinski: Ashlie, this is Cheri. Tom got cut out of the conference call and he's on hold on the conference line...

Ashlie Wilbon: Oh okay.

Cheri Zielinski: ...waiting for an operator. Is there something...

Ashlie Wilbon: Oh no.

Cheri Zielinski: ...we can do here to get him...

Ashlie Wilbon: Yes. (Robert), is there a way that you could reach out to see if he can get reconnected to the call? Are you there?

Operator: I am here.

Ashlie Wilbon: Yes.

Operator: Yes, we're getting people, getting incoming lines. We'll get him on ASAP.

Ashlie Wilbon: Okay. Thank you.

Operator: You're welcome.

Female: I was just trying to recap what I thought was brought up very appropriately by one of the cardiologists, I think, that was...

Male: I remember that discussion.

Female: Yes.

Male: Yes. It's sort of a - the measure was not entirely clear about the systolic and diastolic. And people today usually use the term "heart failure" rather than "congestive heart failure."

Ashlie Wilbon: Yes, sir, that's kind of the way it went.

William Weintraub: I remember the other problem we had with this measure was I think this is the one - it's a year. But it wasn't clear when - how it would relate to a hospitalization.

Ashlie Wilbon: Yes, is Tom there? Or I don't know, Cheri, if you can pitch hit or are you...

Male: Thomas - tom was the one that was commenting on that.

Ashlie Wilbon: Oh.

Cheri Zielinski: Yes. He's still trying to get in. And I don't know what else to do.

Ashlie Wilbon: Okay. We can - we'll just take notes on some of the questions. And when he gets in, we'll fill them back to him.

Jeptha Curtis: You know, Ashlie, this is Jeptha. It took me about 12 minutes to get back on.

Ashlie Wilbon: Oh my gosh. Okay.

Female: And he's on hold for like 5, 6 minutes. So I don't know what else...

Thomas Marwick: I'm here. I'm sorry.

Ashlie Wilbon: Okay. No, I'm sorry, Tom. I didn't realize you got cut out.

Thomas Marwick: Technology.

Ashlie Wilbon: So yes, it sounds like more than one person got cut out. So I apologize for that.

Thomas Marwick: It's okay. All right.

Ashlie Wilbon: So, Kay, can you recap your discussion for Tom since he just got back on the call?

Dr. Kay Reeder: Right. Tom, this is Kay Reeder at Kansas University.

Thomas Marwick: Kay.

Dr. Kay Reeder: We're on the scientific acceptability and measure properties. And Ashlie was recapping and I reiterated on the fourth bullet point that there was discussion and one of the cardiologists in the group had asked too about this, about the use of the terminology "congestive heart failure" as much of heart failure today is just termed "heart failure." And then the ICD9 codes that were included on this or not included on it were discussed relative to CHF terminology. And the physician was wanting to know the delineation of systolic and diastolic.

And I believe you had said that this particular measure was for systolic heart failure. However, if there was a combined systolic and diastolic heart failure present, that it would be in this measure and that Ingenix also has a separate measure for diastolic failure.

Is this correct?

Thomas Marwick: That's correct. And not only is it diastolic. When the ICD9 exists for systolic and diastolic, it's also - it's a marker for severity adjustment.

Dr. Kay Reeder: On your one through four scale that's in the Excel spreadsheet?

Thomas Marwick: No. It's should be in a Condition Status Factor.

Dr. Kay Reeder: Oh. Yes, right. Right.

Thomas Marwick: Yes.

Dr. Kay Reeder: Got it. Okay. Thank you.

Thomas Marwick: Yes. But what you said is exactly right.

Dr. Kay Reeder: Okay. Thank you. So that was - that piece of discussion, that kind of got - didn't - maybe didn't have as much time to discuss as the panel needed. And maybe we want to discuss this further today.

Ashlie, do you think now or do you want to go through...

Ashlie Wilbon: Yes. No, absolutely. I think this kind of speaks to the - I can bring up the criteria. But to me, it sounds like it's a discussion of how precise the specifications are and if what is included in there you think should be included in there and if the way that they delineated and combined the systolic and diastolic is appropriate. So absolutely.



Dr. Kay Reeder: And just as an aside, you may want to take a note. The ABM will need to do that as well. They've got all - everything is combined in there right now. So I don't know if they want - if that's part of what they're going to be looking at in their further study as well.

Ashlie Wilbon: Okay. Thank you.

Male: Well, first, I know there's no good way of delineating between systolic and diastolic failure to administrative codes in any sort of reliable way. I mean - and I didn't see a lot of support for the specific codes that they're using to identify, I guess, as their primary anchor codes that it will get the people into this measure.

And I was kind of wondering or wanted an explanation of the - or maybe rationalization for why they excluded some of the codes that are included in other heart failure quality metrics.

Thomas Marwick: Well again, I think, you know, what we've tried to do was exclude the codes that were specifically for diastolic heart failure. And we basically have another, you know, another episode that we create out of those ICD9 codes. I believe those are the only ones that are excluded. And I think, you know, the attempt here is to try to create a congestive heart failure episode. So that's why we excluded those codes.

We did again include codes that were both systolic and diastolic heart failure. And we took the ones that were both systolic and diastolic and used them as a marker that increases the severity score for the episode. So that's why we did what we did.

Ashlie Wilbon: Dr. Weintraub, I think you had a question, if I have the voices right, while Tom was - had gotten knocked out the call about how hospitalizations were handled. Do you want to...

William Weintraub: Right. So...

Ashlie Wilbon: ...re-ask that?

William Weintraub: ...as I remember the discussion about this before, there's a concern about the hospitalizations when they were occurred during the course of the measure. I can't remember the details of that. But, Tom, you were discussing it the last time.

Thomas Marwick: Oh at the last meeting?

William Weintraub: Yes.

Thomas Marwick: I wish my memory is as good as yours.

Yes, so admissions that occur during the episode that are coded for congestive heart failure, we include in the measure. I don't know if that's...

William Weintraub: Yes.

Thomas Marwick: ...what you're looking for. They would be included in the costs and the claims that grouped to that episode.

William Weintraub: Okay.

Thomas Marwick: And we don't severity adjust for the existence of hospitalization.

Dr. Kay Reeder: Was there - this is Kay again. Was there, in my recollection -- you can correct me if I didn't remember right about the hospitalizations -- something about the timing of those in relation to the measurement period.

Male: And that was a concern.

Dr. Kay Reeder: Right. That's how I recalled that discussion.

Thomas Marwick: Oh. Yes. And I believe that the rule is that if the admission date occurs during the measurement year, then the admission is included in that measurement year. That's the approach we took.

Dr. Kay Reeder: Right. That was more of that, if the admission date occurred during the measurement year.

Thomas Marwick: Right.

Female: And I think if they had an admission for heart failure in the year prior to the measurement year that the patient or the episode was not included. Right?

Dr. Kay Reeder: I think - oh go ahead. I'm sorry. I didn't mean to interrupt.

Female: No, no, I was just...

Dr. Kay Reeder: I - as I recall -- you could tell me if it was wrong in my memory -- wasn't it within six months, if they had a hospitalization within six months prior to the measurement year?

Male: I think that might have been the ABMS measure. I don't think...

Thomas Marwick: Yes.

Male: ...the Ingenix has that particular restriction.

Dr. Kay Reeder: Sorry.

Male: And I don't...

Thomas Marwick: That's okay.

Male: ...think - Tom, remind me if I'm - I think - if I'm correct. Your episodes are insulated, right? They don't consider events that happen before or after the episode?

Thomas Marwick: That's correct.

Male: Right. Okay. So...

Thomas Marwick: They basically create like as much data as you have. It creates a congestive heart failure episode and then it divides it up into years and looks at what's happening during that time. That's correct.

Male: Right. So, Bill, getting to your point, I think that, you know, as we've discussed before, there is the possibility that you're adjusting away for - or you're using as risk adjustment, things that are happening during the measurement year or measurement of the episode.

William Weintraub: Yes, I remember that discussion. So...

Male: Right. That being said, you know, when we talked about this on the steering committee meeting, you know, the predominant sentiment that I got back was that they didn't see that as being a fatal flaw for resource measurement.

Dr. Kay Reeder: Is there anymore discussion on this?

William Weintraub: So it depends on what you're trying to do. If you're trying to predict in the future and - or to correct for that, you wouldn't want to use anything that occurs after the measure starts. You'd want to correct for baseline differences.

If you include things that are occurring in your prediction, there's a circularity. So I'm surprised that the steering committee wasn't troubled.

Thomas Marwick: Well, I mean I don't - I think there's a circularity when what you're using in the current year is some kind of treatment or utilization. But I don't think it's circular if what you're using as a marker is a diagnosis. And that's what we're using, our diagnoses. These are the - this is a...

William Weintraub: I don't know.

Thomas Marwick: ...diagnostic situation that...

William Weintraub: ((inaudible)) I just don't understand, I think, that if you want to risk - you want to risk adjust, if we put people in a level playing field, but in your risk adjustment, you're including something that happens downstream, I just don't understand how that can at all work.

Thomas Marwick: I mean I think, you know...

(Crosstalk)

Thomas Marwick: ...that we're talking about risk adjustment...

William Weintraub: Go ahead.

Thomas Marwick: What we're trying to do is not necessarily predicting the future. What we're trying to do is explain the cost of the concurrent year.

William Weintraub: All right. So over there, there's no adjustment. I mean if you just want to make a measure that says we're going to total the cost, of course you would include it. But if you're going to say what we want to do is level the playing field. These guys are sicker than those guys. So there, we're going to risk adjust for Hospital B being sicker than Hospital A.

Thomas Marwick: Well - but I guess, you know, I don't - this doesn't need to be an argument. I mean, you know, I think it's been set out. But our view is that if they have diastolic heart failure in the year in which we're measuring, that that should be a factor that explains why that particular year costs more money. That's our point.

William Weintraub: You know, but that's not the point. Obviously, if you have a hospitalization, it's going to cost more money.

Thomas Marwick: It's not a hospitalization. It's not a utilization. And I know you have a point about it ((inaudible)) before. And I, you know, I understand that. We've had this discussion before. I don't agree with it, but I understand that. And the - it's not a utilization. It's a diagnosis code of...

William Weintraub: You're trying to...

(Crosstalk)

William Weintraub: ...the money you're going to spend.

Thomas Marwick: I'm sorry. I was talking over part of that. I didn't hear it.

William Weintraub: You're trying to figure how much money you're going to spend, right? The idea is how much money you're going to spend.

Thomas Marwick: Right.

William Weintraub: And you want to show as a measure of how well people are doing that they're spending less money.

Thomas Marwick: Less money given the complexity of the case you're dealing with.

William Weintraub: Right. So - but if the idea would be that you - if someone is hospitalized, you of course, you're spending more money. You're spending more...

(Crosstalk)

William Weintraub: ...- so that's one of the things you're trying to say that those people are spending more money. But then if you go by - back and risk adjust on that, you're risk adjusting the difference you're trying to measure. It just doesn't make any sense. It's absolutely circular.

Thomas Marwick: Wait a minute. I - if you were doing that based on the fact that there was a hospitalization, you're absolutely right. But that's not what we're doing. We're using a diagnosis code that occurs somewhere during that yearlong episode.

Male: So, Tom, I think - not to interrupt, but I think both - like you said, both sides are pretty well laid out.

I think...

Thomas Marwick: Okay.

Male: ...I kind of agree with you that it's not so much that if you risk adjust it for the fact that they were hospitalized during the year, that would be completely circular. I think there is a concern, at least in my mind, that if you are in a - if you're resource - high resource use, you'll have more diagnoses in that measurement period. And that all affect how your - what level of severities you're categorized as being; not so much just because there'll be more opportunities for more diagnoses, is my idea.

Thomas Marwick: No, I think that's a fair concern. But at the same time, you're going to make it, so - and you already have limited ways to severity adjust this. You can make it, but you can't severity adjust it.

And I think, you know, that - and it's not that there are three of these diagnosis codes. It just takes, you know, a single clinical diagnostic marker, not a utilization marker. Yes, you have to have an office visit at least to have a clinical marker, but - so that's the only extent to which that, you know, utilization plays into this.

And I understand the difference of opinion, and I respect that. I just want to make it clear to the committee that we're not using utilization directly to risk adjust the cost of the episode and we are in agreement that that's not the right thing to do.

Male: Right. Okay.



Dr. Kay Reeder: With that being said, this is Kay Reeder again, that maybe there could be more detail and clarification in the risk-adjusted model and the statistical variables used in developing those models, such as your correlations, regressions and so on. I would think we would benefit from seeing some of those things.

Is there - did I say that right or is there anymore discussion on this?

Male: That was one of the things that I - what I thought there was a need for more information. Because one thing that is important in comparing observed to expected is how well the model calibrates. If you can't just really discriminate between the high spenders from the low spenders, then the risk adjustment won't be effective.

Dr. Kay Reeder: Right. Thank you. Anything else, Ashlie, on that?

Ashlie Wilbon: I don't think so. I was just going to try to, since you guys are - we want you guys to try to rate the - each of the sub-criteria while we're on the phone. If we could just kind of refer to - for those of you that are on the Webinar, can you see the - that table that we've been referring to with the side-by-side on there?

Male: I don't see it. I'm not sure; were we supposed to open up that file at the - that was at first slide?

Ashlie Wilbon: Well, no. I guess maybe I'm not sharing my screen correctly. But I just want to kind of go through each of the sub-criteria and just make sure everyone was comfortable and there wasn't any additional discussion on those before you submitted your ratings. So for 2A1 in scientific acceptability, it asks you to rate how well they specified the measure, if it's been precisely specified. So that it could be implemented consistently across organizations. So if two different people picked up the same specifications, would it be - assuming the same data and all that stuff, would it be - could they implement it the same way based on the way the measure specified?

And I think you guys have already addressed that somewhat. But if you have any more discussion on that, we can have that now.

Male: Now this is my impression that it seems to me that the measure is incredibly well specified. It's in - and there's a lot of experience and effort that's gone into their anchoring and their, you know, the specificities or indirectness of the subsequent diagnoses. So, you know, I think it's certainly reproducible if the software is available and if health plans or others want to implement in their populations.

You can argue about the individual codes and the risk adjustment. But I think it is well specified.

Ashlie Wilbon: Does anyone else have anything to add to that or expand upon?

Female: No, never mind. That's 2B that I'm looking at. Never mind. Go ahead.

Ashlie Wilbon: Okay. And we can move on. So the next one - and for those of you that are able to open the SurveyMonkey link, as we're going through these, as you hear the discussion, if you're ready to submit your ratings, you can go ahead and do so.

2A2 focuses on whether or not the reliability testing has demonstrated that the results are repeatable and would produce the same results, a high proportion at the time. So I'll open that up for discussion.

Male: Can I just ask you? I didn't open up the SurveyMonkey at the beginning. Is there any way to get back to it so that I can open it up?

Male: Do you have a link in Ashlie's e-mail?

Male: Yes. Yes, that's Ashlie's e-mail from the 11th, 5:36 pm.

Male: I'll look for it. Thank you.

Ashlie Wilbon: Yes. So perhaps, Carlos, if you want to - I think you might have already mentioned this.

But I might have stopped you before you got into reliability. But if you want to recap your assessment of reliability for the group?

Carlos Alzola: Sure. In terms of the reliability here, we're looking at repeatability of the measure. And we know that we can do this by either looking at the (real possibility of the data or (real possibility) of this call itself. The way they did it is by performing parallel development of the data and this call by using two independent approaches -- one, using the ATG software and another one using a soft prototype.

So with the point into - these two independent approaches would lead to the same results. And so they defined agreement between the two approaches as an exact match between the grouping of records and an assignment of resource use. And their matching rate that they obtained was over 99.9%. So in my view, that proves reliability very strongly.

Female: It was for the samples that they used, keeping in mind that these are commercial data and I think that the age limit on the - for adults was 65. So they did not have people included who are over 65.

Carlos Alzola: I think the database does include people over 65.

Female: Oh it does?

Carlos Alzola: Yes, but I...

Female: Oh.

Carlos Alzola: ...think it's being considered for people over 65. It's not being considered for approval for people over 65.

Female: Oh. Right. Okay. Thank you.

Male: That's a good question. So, Ashlie or Tom, can you just clarify that because it does impact a little bit about this?

Ashlie Wilbon: Yes. I'm actually just trying to reconcile that in my head.

So maybe, Tom, so can you just clarify the testing that was done, what was in the testing database that you guys did, the age limit?

Thomas Marwick: Yes, I believe it was mostly a commercial database. We had some Medicare patients in that population. They're not really - they're - you know, they like that style. I think they're Part, like Part C plan Medicare patients. So the majority of the patients were commercial and under the age of 65. And we're seeking approval for commercial use.

Female: Right. So if there are patients over 65 in the commercial data that they would be okay or am I not...

(Crosstalk)

Female: ...interrupting that correctly? Maybe...

Male: Yes.

Thomas Marwick: I think basically it doesn't matter because they're tested within the commercial population. They're seeking approval for it in the same population.

Female: Okay. Okay, thank you.

Female: Are we ready to move on then?

Ashlie Wilbon: Sure. So it sounds like reliability testing is out of the way. We'll move on at 2B1, which looks at actually the rate, whether or not the measure specifications are consistent with the focus of measurement and the intent of the measure as outlined in the Importance of Criteria 1B, and that it specified to capture the most inclusive target population indicated by the evidence.

So this again is more so about the clinical construction and whether or not it's measuring what they say it should be measuring.

Dr. Kay Reeder: And it's a measure for systolic or combined systolic/diastolic heart failure. Is that right?

Male: I think that's correct. Right.

Thomas Marwick: Yes.

Male: So again...

Thomas Marwick: If you're asking me.

Male: ...I think we've discussed this. But it seems like it's - that's one of the things that's controversial, is the completeness of that or the accuracy of that assessment, like can you focus on just that population. But I think we've, at least in my mind, discussed it. But we're certainly open to additional questions.

Dr. Kay Reeder: Right. It was just - I just wanted to bring it up that that was where we were with that piece of it. But I don't have any more discussion on it.

Does anyone else have more for that one? Can we move to 2B2, Ashlie?

Ashlie Wilbon: Yes. So 2B2 focuses on the validity testing of whether or not the testing has demonstrated that the data elements or the measure score correctly reflects the cost of care of the resources being used. And I don't know if Carlos wants to maybe recap some of his assessment of the validity?

Carlos Alzola: Sure. Again, for validity, they look at face validity. And they look at the face validity of the measure score. Through that, they look at - they provide some tables, looking at the cost and utilization measures for different lines of service. And they do that across the different severity levels for different peer group definitions.

The largest component of cost is the hospital expenses and then the specialist's cost per episode. And the - this cost is positively correlated with severity. So that seemed to be - show the face validity of the measure.

Dr. Kay Reeder: Thank you. Is there any further discussion on that point?

Ashlie, can we move to 2B.3?

Ashlie Wilbon: Sure.

Dr. Kay Reeder: On the exclusion, there's just one thing with the notes. I need clarification on the discussion from last time.

Ashlie Wilbon: Okay.

Dr. Kay Reeder: I'm not on the screen, but I found my handout.

Ashlie Wilbon: Oh okay.

Dr. Kay Reeder: Under discussion, it says the TAP suggested including codes for 28 and for 28.1. And I guess at this point, because of my Excel spreadsheet includes those codes, Tom, were they not included? Were they excluded before or - because I show that they were already there.

It's just a point of clarification because from the last phone conference discussion it says that the TAP suggested including codes for 28 and 428.1, but according to the Ingenix Data Dictionary Excel Spreadsheet, I show that those are there already.

Ashlie Wilbon: Right, I'm not sure if (Tom) is still there, but...

Dr. Kay Reeder: Oh.

Ashlie Wilbon: I'm not even sure exactly -- does anyone know what the 428 is?

Thomas Lynn: I was on, I'm sorry.

Ashlie Wilbon: I know that was pulled out of the transcript so...

Thomas Lynn:: I'm on you, you're right, they are included.

Dr. Kay Reeder: 428 is heart failure, 428.1 is left heart failure.

Thomas Lynn: Left heart failure.

Dr. Kay Reeder: Right and I show that they are included so somewhere along the line either the transcript has an error in the transcription or we didn't realize it was already included in our last discussion. Would that be fair?

Thomas Lynn: Honestly, if my memory serves me, I think the TAP suggested we remove these codes.

Dr. Kay Reeder: Oh, remove them.

Thomas Lynn: Yes. And I think it's because the argument was that they maybe weren't specifically systolic or congested in nature. 428 is a miscode anyway, meaning that, you know, in certain situations it would not be -- it wouldn't be -- you wouldn't pay based on it so people don't use it very often. 428.1, left heart failure, you know, could be used -- and I think the person on the call was arguing that maybe that didn't really represent congestive heart failures specifically. But it was -- I believe the argument was to remove those two codes, which, you know, we can easily do, but...

Dr. Kay Reeder: Yeah.

Jeptha Curtis: That's closer to what I remember. I think it was Dr. (O'Toole's) concern from the last call.

Dr. Kay Reeder: Right. What can we -- is that the recommendation today?



Jeptha Curtis: I don't know, what do you recommend?

Dr. Kay Reeder: I would have to defer on this one to the -- I don't have a recommendation; I defer and go with the TAP on this one.

Thomas Lynn: I'm glad to hear you guys are struggling with it too.

Dr. Kay Reeder: I do remember the conversation but I didn't remember the conversation matching what is written as a discussion point. So that's why I say something was lost in translation/transcription or I heard it wrong last time or whatever. But it is supposed to or included currently, so I think there's a mismatch in our -- in what either was discussed or transcribed from our discussion and I really don't have a recommendation. I think we just need to clarify it.

Jeptha Curtis: Right. I don't think there's a right or wrong way to do this. You know, taking the ((inaudible)) and they're trying to focus on systolic, that's not a code that's consistent, absolutely consistent with systolic but, you know, they made a judgment call. I'll just restate that I think trying to distinguish between systolic and diastolic is difficult and so like most of the quality measures do, I'd recommend lumping them all together for one giant heart failure measure but, again, we've talked about that.

Thomas Lynn: Right and, yes, you know, right, we're just trying to make a judgment call, we made judgment calls to include those two claims, I mean ICD9 codes and this, we made a judgment call, actually, well I think it was pretty soon after they created the diastolic ICD9 codes to pull them out into a separate episode, you know, that's our judgment call. I mean it's not, we didn't make it in and the ((inaudible)) is certainly consulted, you know, cardiologists and our advisory board before doing that.

Dr. Kay Reeder: Okay. Is there anything else on exclusions and on the 2B point 3?

Jeptha Curtis: I mean there really are no exclusions, right, its only data exclusions, not patient exclusions?

Dr. Kay Reeder: Correct.

Ashlie Wilbon: So we can move onto 2B4 it sounds like. 2B4 is about the risk adjustment model and whether or not the strategy that they have described is based on patient clinical factors that impact the measured outcome -- or the measure outcome or the resources and that our present at the start of care and that they've demonstrated adequate discrimination and calibration.

Dr. Kay Reeder: I would like to ask the statistician to comment on this.

Carlos Alzola: Sure. The comment is the same, I have had before and that there is lack of information in terms of the includeness of feed and the calibration. The risk groups are selected in terms of some cutoff for the severity score and I don't have any information on to why those particular values were chosen. As for the other question, I should comment on the other question of whether including the diagnosis that occur after the patient as selected for the population, I think its fair to include those things in the model and the main reason is that since we want to make comparisons fair, we need to know that a patient has a condition that makes it more expensive to treat.

So we're not evaluating whether he got the right treatment or not but if a patient has some additional ((inaudible)) that make it more expensive to treat, then it would be unfair, in my opinion, to not consider that because if -- only because it happened after they were selected because then the patients that did not get sicker would be -- would appear as they received more efficient treatment, they were treated more efficiently.

But I mean the resource use was used more efficiently, but we really don't know that.

(Sherry): Hey Ashlie this is (Sherry) I just want to let you know that Tom got cut off again so I'm going to have to take notes until we get him back on.

Ashlie Wilbon: Okay, thank you. Sorry about that.

Dr. Kay Reeder: Is there any discussion on that?

Jeptha Curtis: From my perspective it's -- we've discussed the risk adjustment and the pros and cons in their particular approach and Bill do you have any other thoughts or...

William Weintraub: No I've already expressed myself on it.

Dr. Kay Reeder: The only thing I have and it may not even be an issue is the winsorization process that they use, they didn't identify the cutoff level and then they only did it on the high end and I put that in my survey monthly from the last -- after the last conference call on that one and like I said it may not be a big issue but it was just me wondering about it.

Carlos Alzola: I have the same question and I tend to think that that's because the low outliers are those are excluded while ((inaudible)) are not and I think it's because the ((inaudible)) are incomplete episodes but they didn't really say it explicitly in the submission I don't think.

Dr. Kay Reeder: Okay, thank you. Is there anything else on this one?

Ashlie Wilbon: Okay so 2B5 looks at whether -- asks whether or not the measure has demonstrated that you can identify statistically significant practically and meaningful differences in performance.

Perhaps Carlos would like to summarize a little bit about his assessment on this based on what was submitted?

Carlos Alzola: Yeah. There is no assessment here in terms of considering -- comparing statistical versus practical significance. I think the way they summarize this clause is that they define/ask the user what clinically significant -- clinically significant difference it would be for them and then use the formulas to determine the competence intervals and that's -- and using those formulas they can -- they can see if there sample size is sufficient enough to obtain a competence interval that will be useful in establishing differences that are clinically and statistically significant. Maybe the developers want to comment on that?

Thomas Lynn: Yeah, what we're doing is we're basically creating competence intervals around the ODE ratio and stating that they're statistically different if they don't cross a threshold. I believe in the case of the sample, the threshold is the expected ratio for the peer group which would be one and if the competence interval crossed that threshold -- crossed one then it couldn't be statistically -- the difference wasn't statistically significant if it didn't and it was. I believe that's how it works.

And what we're trying to do is take out of play the count - the minimum count, which, you know, may vary depending upon the case mix of the physician and how different they are from the expected value. So we feel like the better thing to do is to, you know, be specific and look for a statistically significant difference, and that's what we try to do.

Carlos Alzola: Okay, but then you don't really make an attempt to determine what is clinically significant, like it's the difference of 1.05 versus one, versus one.

Thomas Lynn: Right, that's not in there, it's just looking at statistically significant differences.

Dr. Kay Reeder: Is there any more discussion on that?

Ashlie Wilbon: Okay so the next sub criteria for scientific acceptability is 2B6, it looks at multiple data sources and again that's one that we've been putting not applicable for because the measure uses admin data and there's no other types of data that have been specified. The last criteria for scientific acceptability addresses disparities and whether or not they've been addressed in the measure based on how they have identified them in the important section.

Dr. Kay Reeder: I don't see any concern from my perspective on the -- on disparities. I mean they've looked at age, gender and race, unless someone else has any discussion on that?

Ashlie Wilbon: I'll just add in that I realize this disparity discussion is one that has kind of over arches a lot of the measure discussions we've had across all the TAPS and the Steering Committee as well and this is something that we've kind of identified that there are some, you know, short comings to admin data and being able to measure this.

But there, perhaps, is some discussion about the ability of the measures current structure to be able to stratify for this if the data was available. So that may be a consideration or if you're satisfied with the way that the measure is currently structured and reporting out or if you feel like disparities are significant enough for this condition that it's something that should be addressed.

Dr. Kay Reeder: And so if it's the latter then we should look at is as an insufficient -- is that what you're saying?

Ashlie Wilbon: Not necessarily but if you have some -- maybe you want to kind of talk about what your concerns may be about that, I was just kind of throwing those out of considerations when you think about this criteria but...

Dr. Kay Reeder: Someone had started to say something, do you want to continue? Was it you Dr.

Curtis?

Jeptha Curtis: I just felt like we have to put -- or I have to put insufficient just because I didn't see any real considerations of disparities or an explanation of why they weren't considered and I think it has to do with limitations of the databases that they had for testing, at least for (Rays). But, you know, I'm not sure exactly how disparities and resource use have been -- will play out in the public reporting realm.

Dr. Kay Reeder: Good points. Okay, Ashlie.

Ashlie Wilbon: Sure, so it looks like we've finished up scientific acceptability and we're a little bit behind. If we can kind of get through usability and feasibility quickly and move onto 1594 that would be great. So usability -- so hopefully those of you that have your online surveys open, if you can kind of just wrap up your online evaluations for scientific acceptability and then we'll move quickly through usability and feasibility and you can submit for that measure.

Dr. Kay Reeder: Ashlie?

Ashlie Wilbon: Yes.

Dr. Kay Reeder: I'm sorry, I'm going to interrupt you, but do we need to be looking at the handout that we had previously for how we had looked at usability and feasibility? It was a table...

Ashlie Wilbon: Oh, I think those were for the ABMS, oh no it was for all three.

Dr. Kay Reeder: It was for all three.

Ashlie Wilbon: It was. I'm not sure, Jephtha do you think that's something we need to -- I know we were using that for internal consistency purposes, I don't know if you have any thoughts on that Jephtha?

Jephtha Curtis: If you have a conveniently available it might be useful but otherwise I think -- I think we should just keep going through it.

Dr. Kay Reeder: Okay.

Jephtha Curtis: But just to set the precedent or the foreshadowing here -- I would love to be done with these measures by the end of today, I think we have to be for the Steering Committee call tomorrow and I'm hoping that people might be willing to stay a few minutes later to just get through the other measure completely.

Ashlie Wilbon: Yes, I would agree, I think we want to be done just as much as you guys, so let's just try to move quickly through the usability and feasibility should go relatively quickly as well. So usability is the first one and asks whether or not the performance measures are reported as to the public at large at a national community reporting program and if there's been any consideration -- I'm sorry, if exceptions may be considered if there's evidence that the measured performance results are available for public reporting and that the measure has benefited the public. So basically we're asking whether or not this information is acceptable to the public currently and again I think we've had this discussion for other developers and for other Ingenix measures in particular and I think, I don't know if Tom if you want to just recap how your data again is --or your measures are used outside of -- in the public or in the community.

Thomas Lynn: (Sherry), do you want to do that or do you want me to?

(Sherry): Yes, I think we've had this discussion as well on other measures as you've said, you know, our measures are used and widely spread to measure, you know, provider work in these areas and to compare the providers to one another, there's provider profiling, provider report cards, there's cost base analysis for the members to estimate what the cost of the service would be if they, you know, if they encounter this episode and you want to know how much out of pocket expenses is going to be. Are products are used for those, you know, they are the basis for risk adjustment for actuaries, you know, it's pretty widespread.

Jeptha Curtis: Right, and just to recap what we discussed before, it's in -- the overall Ingenix measures are in wide use, this particular measure has never been broken out to it's usability or whether or not it specifically has been able to be incorporated in QI programs. Right.

Dr. Kay Reeder: That is correct.

Thomas Lynn: We use it in combination with other measures, that's correct.

Jeptha Curtis: Does anyone have any questions about the usability of this particular measure?

Ashlie Wilbon: So the next criteria of 3V asks whether or not the performance results are meaningful and understandable to the intended audiences. And then the next criteria after that I'll just go ahead and read through them, ask about whether or not the measure - clinical construction logic - is transparent and can be broken down and understood in that way.

Dr. Kay Reeder: Thank you Ashlie.

Ashlie Wilbon: Any discussion on those items?

Jeptha Curtis: I can speak for...



Ashlie Wilbon: I am actually going to try to bring up some of the reading results for the old and the other Ingenix measures.

Jeptha Curtis: If there's nothing different on the usability and feasibility then what was previously voted on for the Ingenix measures, would be...

Ashlie Wilbon: Right. I would agree. So just quickly to recap, 4A and 4B for feasibility, generally we don't spend time on it, asked whether the data available electronically and for be looked at whether or not -- I'm sorry, whether the data elements are routinely generated and whether or not they're available electronically is admin data, so by default we've kind of been skipping over those. 4C addresses whether or not the measure or implementing the measure that there is a lot of susceptibilities to inaccuracies or errors and unintended consequences of whether or not they can be monitored or detected.

4D looks at whether or not a data collection and measurement strategy can be implemented or if there's any barriers to implementing the measure.

So in the interest of time, unless anyone has anything new to discuss on these as we have discussed with other Ingenix measures, I would just ask that you submit your ratings online. If you have any specific rationale to submit with those ratings that you would do so in the online tool and we'll capture those and then you can submit that now if you can.

Dr. Kay Reeder: Ashlie?

Ashlie Wilbon: Yes?

Dr. Kay Reeder: I think I've already submitted this one.

Ashlie Wilbon: Okay, that's fine, we can pull...

Dr. Kay Reeder: From the June 9 is -- you don't need it again do you?

Ashlie Wilbon: No. As long as you don't feel like you would change anything, we can use that rating.

Dr. Kay Reeder: Yes, yes. Okay, thank you. Thank you everyone.

Ashlie Wilbon: So is everyone okay with moving onto 1594 at this point?

Thomas Lynn: Absolutely.

Ashlie Wilbon: Okay. Let's transition over to 1594 and if we can start off again with develop we're doing, just a brief introduction, this will be the first time the TAP will be discussing it and we'll go on from there.

Thomas Lynn: Yes, this is -- I'll be very brief because I know we're running a little behind but this is a very similar idea. It's creating an episode of coronary artery disease using claims administrative data, dividing those -- that into year-long episodes and then I'm using diagnosis based clinical markers to do severity adjustment of those episodes and I'll leave it at that.

Jeptha Curtis: Tom this is Jeptha. Can you just expand on how differed or overlapped with the AMI measure that you had us consider earlier?

Thomas Lynn: You're asking me that?

Jeptha Curtis: Well I know the difference between them but I think there is some overlap in terms of how the patients that are entering it, maybe I'm right.

Thomas Lynn: I just couldn't answer that question. I'm not familiar with the ABMS measuring in detail.

Jeptha Curtis: I'm sorry, I meant the Ingenix measure for -- you had an AMI Ingenix measure, correct?

Thomas Lynn: Yeah, you know, that measure was -- this is a very different measure in that it's looking at a chronic disease, a year-long episode of the chronic disease. The AMI measure should have been developed as an event and it was not, it was -- in a lot of ways it was similar, what we submitted was similar but it was less on target because that should have been an event not looked at as a disease or condition. But we did, in fact, for that measure use a coronary artery disease year-long episode that included an acute myocardial infarction. So this would be that same thing except we would use all of the episodes of coronary artery disease, not just the ones with acute myocardial infarction.

Jeptha Curtis: Thanks. Ashlie, who was the primary reviewer on this?

Ashlie Wilbon: Right so Dr. (Marwick) and Dr. (Palastrant) were primary reviewers and neither of them were able to attend -- I thought Dr. (Palastrant) was going to be on the call but he sent me an email and said that something had come up so we don't have either of the primary reviewers on the phone but we can try to use our resources through Carlos and to get some additional input on it.

Jeptha Curtis: I mean I've reviewed it in a fair amount of detail so it...

Ashlie Wilbon: Okay great.

Jeptha Curtis: For the people on the -- who haven't had that opportunity and I think the methodology as Tom said is almost identical to the other resource used in that. Sorry, there's a lot of motorcycles going by. The methodology is very similar, it's really just a matter of how they're identifying patients in the -- to go into this episode and a little bit about the risk adjustment were kind of the two biggest differences that I think we would probably be well served focusing on.

The -- I guess the question that I had initially when I was going through it and I'm trying to pull it up. Ashlie can you pull up the excel spreadsheet that had the primary diagnosis codes on it?

Ashlie Wilbon: Sure. Do you know which one? Was it labeled...

Jeptha Curtis: I think it's S5 CAD Data Dictionary...

Ashlie Wilbon: Five?

Jeptha Curtis: Yeah, primary diagnosis codes.

Dr. Kay Reeder: Yes, that's it.

Jeptha Curtis: Right. So, you know, I don't know if people can pull it up or look at it, but, you know, they have, you know, the diagnosis codes are basically the 410's through 414's and then the 429's, which represent complications of MI. So there's, I think, my question was that seems comprehensive in some respects for identifying patients with coronary artery disease. It does raise the same question that we touched on with the AMI measure as to how comparable are these populations. Are they similar enough that you can reasonably make inferences about the resource use in someone who just had an MI or someone who had their MI or their PCI several years ago and still carries the diagnosis.

Tom I don't know if that's making it clear but I think that's the key thing that needed to be rationalized that I didn't really see in the application.

Thomas Lynn: Yeah I think that's right and I think, you know, there was an effort made to -- a decision for us on this rule was to, you know, build in severity adjustment for those sorts of things and that's what we did. But I can see, you know, I can see where this is different than CHF because it maybe sort of a broader -- we're starting out with a wider distribution. I don't know if I have -- I don't know if we have evidence that it's really wider distribution or not but I hear what you're saying.

Jeptha Curtis: Other members of the TAP, do you have any thoughts on that?

Male: Yeah, just briefly, one question, if someone has CHF -- excuse me, CAD, as listed as having CAD, are they considered to continue to have CAD indefinitely?

Jeptha Curtis: I think it's similar to what they have for their other measures. If they have another CAD code or related code within a certain amount of time, the episode continues up to a year, is that right...

Thomas Lynn: It's exactly the same as CHF, that is correct.

Jeptha Curtis: Right. So I think it's probably, and my expectation would be that most patients would carry this as a chronic disease and not, you know, be a limited episodes within a year.

Male: Yeah but if CAD code is mistakenly dropped, does that create problems?

Thomas Lynn: Well what it causes is that, you know, if it was really dropped for over a year, then there would be a year where there wouldn't be any cost and there wouldn't be an episode.

Male: So as long as it shows up for once in the course of a year.

Thomas Lynn: Right.

Male: Okay.

Jeptha Curtis: So, then, so, you know, in terms of the completeness of it I think -- I didn't find anything that seemed to be really missing in terms of diagnosing coronary disease or, you know, missing big chunks of the population similar to the heart failure measure there's not a lot of medical exclusions or patient exclusions -- there simply data exclusion criteria that are -- well I think are all identical to the ones that we've reviewed for the other measures.

Thomas Lynn: That's correct.

Jeptha Curtis: And then -- I'm trying to go down to the, you know, the attribution, no, not the attribution, but the risk adjustment -- what page of the PDF is that on, do you know, of the clinical severity levels?

Ashlie Wilbon: One second, let me bring it up -- 22.

Jeptha Curtis: Okay, I'm having a hard time finding my notes on that. But, again, it's the same methodology but, you know, that you are accounting for, comorbidities and disease severity based on episodes that happened within that year that are deemed related to the CAD diagnosis and creating a severity score which then, I think, groups into, is it still five groups or four groups total?

Thomas Lynn: It's four groups and you're accurate in saying that it's -- that we do it the same way and look at things that occurred during the episode.

Jeptha Curtis: Right. I guess the only thing that I would have liked to have seen is kind of how that, well number one it had the question is that enough groups for something where you can have someone with an acute MI with cardiogenic shock needing an Elveden transplant or something like that and the same measure as someone who had a PCI five years ago and still carries that diagnosis. So again, just it emphasized that example for me the extremes of resource avidity that you would have in those two patient populations.

And I'm trying to think of what else is different here or pretty much everything else was really the same. We didn't, I think, talk too much about the physician attribution issues within the heart failure measure but, again, they're the same here. But it's similar in that they don't specify a specific approach to attribution but they do specify different -- or they outline different approaches that a user could take to defining attribution. So it's a little difficult to say, this is one thing the Steering Committee struggled with a little bit on -- if it's not defined as to how to attribute yet it could be attributed to the individual physicians, can we give this the stamp of approval as being NQF approved.

Thomas Lynn: Well what do they say?

Jeptha Curtis: I'll say that the voting was very mixed and with strong passions on both sides, I think that's a fair assessment, actually is that fair?

Ashlie Wilbon: Yeah, I would agree, I would agree.

Jeptha Curtis: So why don't we start going through this scientific acceptability point-by-point then, that was kind of my overview of kind of what I thought the important things to consider were.

Ashlie Wilbon: Okay.

Jeptha Curtis: And just go through it and see if there are things that people want to raise.

Ashlie Wilbon: Sure. So 2A1 again is about how well the specifications were written and whether or not based on how they're written now would they be able to be implemented consistently across organization based, you know, assuming all is equal, were just based on the instructions that are written in the specifications.

Jeptha Curtis: Right and I'd have to say again like I did for the other measure, that it's very well specified, they've already shown that they can specify the measure across different health plans and are doing this currently.

Ashlie Wilbon: And so the next one for 282 is on reliability testing. I would just ask -- Carlos is there anything different about this measures reliability testing, I'm assuming not, but I just -- have you identified anything that was different from this measure that was not identified in your previous assessment for 1591?

Is Carlos there?

Carlos Alzola: Hello.

Ashlie Wilbon: Hi, Carlos?

Carlos Alzola: Yes.

Ashlie Wilbon: Okay.



Carlos Alzola: No, it's -- they use exactly the same process, there really wasn't any difference.

Ashlie Wilbon: Okay, okay, thank you. Any discussion on the reliability? Okay, for 2B1, which is the first criteria on validity, it asks whether or not the specifications are consistent with the intent -- with the measures intent and the focus area that they're on, so are they measuring what they say they are measuring.

Jeptha Curtis: And I guess it comes down to the -- I'm a little mixed on it by I think overall they do define a population that they say they're going to capture and they capture it. I think most of my concerns come down a little bit later in terms of the risk adjustment but, you know, I thought that, again, there's not a lot of exclusions that have been identified other than the data exclusions so you can't really comment on whether or not they're supported by the evidence. But I don't think that they were missing a large part of their target population.

Ashlie Wilbon: Okay. And 2B2 is on validity testing, again, I'll just ask Carlos, is there anything different about the -- your analysis of this measure in comparison to the other Ingenix measure?

Carlos Alzola: They used the same methodology. They look at cross manipulation for ((inaudible)) different severity groups and they just verify the dose were what they would expected clinical.

Ashlie Wilbon: Okay, and then 2B3 is focused on whether or not, and I think you've mentioned this already, whether or not the exclusions are supported by evidence and that the measure specification are included -- where they've identified exclusions that they're transparent and I think we've discussed that already so I'll move onto 2B4 where we addressed risk adjustment and whether or not their risk adjustment strategy is based on patient clinical factors that influenced resource use and that are present to start a care and whether or not they have demonstrated adequate discrimination and calibration.

Jeptha Curtis: So that's where I'm going to have to ask Tom again, is there any place in your application you can kind of point me to that shows how well this accreditation of severity works. Like in a, you know, just try to convince me that it's worth incorporating all this population that you're adjusting for the differences in a fair and equitable manner?

Thomas Lynn: Yes, so what we -- we didn't do that with this application actually, we were asked to do it with diabetes and we did, we basically, you know, did a correlation with the severity levels and the cost that the episode ending up requiring. And (Sherry) did we end up doing that for the other episodes too or did we just do it for diabetes, are you still on the line?

(Sherry): I am and I don't recall off the top of my head, I'd have to go back and look.

Thomas Lynn: But we could -- that's something that we could send along if you thought that would help.

Jeptha Curtis: Yeah, I think it's going to be tough though because if this Steering Committee's meeting on it tomorrow, I mean...

Thomas Lynn: I know.

Jeptha Curtis: ...the only thing, I certainly wouldn't expect you to be able to run it by tomorrow, but maybe having the diabetes one available to kind of talk people through. I don't know actually if we could have it for the Steering Committee...

(Sherry): Yeah actually...

Jeptha Curtis: ...if they ask for it?

(Sherry): ...we're wrapping up the diabetes measure tomorrow with the Steering Committee, so it wouldn't be too farfetched to kind of say -- I mean I guess it would be the numbers wouldn't be exactly the same but the process that they use would be similar I would assume.

Thomas Lynn: The process would be the same but I think you're argument that there's a difference here in coronary artery disease because of the wide disparity of care.

(Sherry): Right.

Thomas Lynn: You know, would not be addressed, but if we haven't done it we won't be able to get it by tomorrow but if we had done it then it might be lying around someplace so if we get it we'll forward it onto you guys for discussion tomorrow, if not you guys are going to either have to make a decision without it or give us a couple of days to send it around. Sorry about that.

Jeptha Curtis: Yeah, no and Ashlie I wonder if, if I recall correctly, then after tomorrow's meeting, or at some point in the next few months, these will go out for public comment and then there will be kind of a retrenchment after that or as we look.

Ashlie Wilbon: So I'm just, I was just thinking aloud with (Terrune) here, I think there may be opportunity, it's hard to say right now because we are kind of -- we're moving relatively quickly, but there may be an opportunity, like you said, after the comment period or, I mean right now I would say that the TAP has to evaluate the measure as is and what's in front of them and we would ask, you know, the Steering Committee to do the same and then at public comment period you would be able to provide the additional information and the Steering Committee can evaluate that and decide whether or not they want to change their vote based on that additional information which should happen so.

Jeptha Curtis: So I think that's fair. I think as is, I'd have to rate this as a low personally, just because I don't think it's a compelling...

Thomas Lynn: Or is sufficient.

Jeptha Curtis: ...argument that it's working, I mean, yes.

Ashlie Wilbon: Right. So Tom we can talk offline or email after the call today and figure out -- I just need to kind of check internally and see what strategies we might have for you guys to be able to do that and what opportunities you would have to do that.

Thomas Lynn: Okay, thank you.

Ashlie Wilbon: Okay, thanks.

Jeptha Curtis: And then for 2B5, data analyses, demonstrate methods for scoring analyses allowed for identification and statistical significant practically meaningful differences and performance. Again, I think this outline is comparable to what we had for heart failure measured in that and Carlos feel free to jump in here. You've identified an approach to doing that, I didn't see a lot of...

Carlos Alzola: No.

Jeptha Curtis: ...examples that demonstrated how that plays out practically.

Carlos Alzola: No, no it was exactly the same, since they only look at clinical significance -- a ((inaudible)) significant difference.

Ashlie Wilbon: Okay, again, 2C, 2B6 would be not applicable and 2C I would assume would be very similar again in the way that -- with the admin data and the availability -- and the ability or capability of the data to report out or capture disparities. So your rating on that would be consistent from the other measure and is there anything else on scientific acceptability before we move on to usability and feasibility?

Kay Reeder: I don't have anything, this is Kay.

Ashlie Wilbon: Okay. And again I think that will be very similar. I did bring up the memo that I sent for our call for June 9 where it captures the ratings for each of the developers in usability and feasibility so that you can kind of see for consistency - I'm going to drag it over to the screen here - for consistency to see how -- how you rated things and it looks like for some of the criteria there were some insufficients, typically for 3A, can, those in the webinar, can you see this...

Jeptha Curtis: Yes.

Ashlie Wilbon: ...with the yellow, gray and blue? Okay, I'll try and make it bigger. So I'll just leave that up and we can discuss with it on the screen or Jeptha I don't know if you have any specific comments about any of these?

Jeptha Curtis: No I mean, again, it's -- the rationalizations are exactly the same, it would be hard to differ too much in the voting and I think it kind of if you look at the prior scores for usability, you know, it's a lot of insufficients and moderates, just because being asked to extrapolate from fairly vague comments which I understand why they couldn't be more specific, but, again, it's the same issues.

Ashlie Wilbon: So if you, again, if you have the link up on your screen and you're able for those TAP members on the call to submit your ratings while we're on the phone, that will be fabulous. If you

have any trouble accessing that link or getting into it, please let me know, I can help you, (Sarah)  
or I can help you get into that and help you get your ratings in. We want to try to turn it around  
again for tomorrow.

Anything else?

Jeptha Curtis: No, it's remarkable how fast we were able to go through it. It's actually, in some ways,  
much easier to go through each developers one after another, as opposed to switching gears  
each time.

Ashlie Wilbon: Yeah, I would agree.

Dr. Kay Reeder: Yeah.

Jeptha Curtis: But I don't want Tom and the folks from Ingenix to think that we're not thinking seriously  
about this.

Thomas Lynn: No, no, no.

Jeptha Curtis: It's just like -- I feel like we have reached our -- it's almost like a political debate. You know  
where you stand on most of these things.

Thomas Lynn: Right and then I think you captured the differences though, I'm okay.

Jeptha Curtis: Okay.

Kay Reeder: No, I appreciate Dr. Curtis the attention you gave to it in pointing out the broad spectrum  
and the concerns within that, so I thought there was a lot of the same type of approaches they

had for the heart failure which I went pretty deeply into, probably deeper than this one but I found very much similarities as you mentioned.

Jeptha Curtis: So actually I think that is pretty much everything, right?

Ashlie Wilbon: Yeah, that sounds great and like I said we'll follow-up with you guys via email on the last -- the two ABMS measures to see if you have any input you'd like to provide us to pass onto ABMS as they work on refining they're measures. We will, for the Ingenix folks, we'll be compiling what was submitted today from the TAP and you'll see it on the call tomorrow with the Steering Committee. Unfortunately we don't have the benefit of being able to do real time voting like we do and the End Person Meeting because the limitations of the tool, but we're going to turn it around so you can see the outcome of the ratings and we'll be recapping the TAP discussion today on the call tomorrow with the Steering Committee.

Thomas Lynn: All right, we'll see you tomorrow.

Ashlie Wilbon: Okay, thanks Ingenix and I wanted to thank, on the record, both (Jamie) and Jeptha for their leadership and there, you know, help in leading the TAP through this effort and we realize it was a herculean effort being the first TAP, having the most measures with the most developers and everything, so I appreciate again your perseverance and all of your time and I realize this is third and fourth job for many of you so I appreciate it a lot.

Kay Reeder: Thank you so much Ashlie.

Ashlie Wilbon: Thanks everyone.

Jeptha Curtis: Yes, Ashlie, great job for you and everybody at NQF and my thanks to everybody on the working group.

Male: I thought the staff did a wonderful job.

Ashlie Wilbon: Thank you guys. Okay take care.

Thomas Lynn: All right, take care.

Male: Bye-bye.

Ashlie Wilbon: Okay.

Kay Reeder: Thank you very much for...

Ashlie Wilbon: Oh, you're welcome.

Kay Reeder: ...enduring all of that. It went well, we got them all done so that they can meet with the...

END