NATIONAL QUALITY FORUM

+ + + + +

RESOURCE USE CARDIOVASCULAR/DIABETES TECHNICAL ADVISORY PANEL MEETING

+ + + + +

WEDNESDAY

MAY 11, 2011

+ + + + +

The Technical Advisory Panel met at the National Quality Forum, Suite 600 North, 601 13th Street, N.W., Washington, D.C., at 8:30 a.m., Jeptha Curtis and James Rosenzweig, Co-Chairs, presiding.

PRESENT:

JEPTHA CURTIS, MD, FACC, Co-Chair, Yale University School of Medicine

JAMES ROSENZWEIG, MD, Co-Chair, Boston Medical Center and Boston University School of Medicine

MARY ANN CLARK, MHA, Neocure Group

CONSTANCE HWANG, MD, MPH, Resolution Health, Inc.

THOMAS MARWICK, MBBS, PhD, Cleveland Clinic

DAVID PALESTRANT, MD, Cedars-Sinai Medical Center*

BRENDA PARKER, PharmD, GlaxoSmithKline

KATHERINE REEDER, PhD, RN, University of Kansas School of Nursing

WILLIAM WEINTRAUB, MD, Christiana Care Health System

NQF STAFF:

TAROON AMIN, MPH

HEIDI BOSSLEY, MSN, MBA

SARAH FANTA

SALLY TURBYVILLE, MA, MS

ASHLIE WILBON, MPH, BSN

ALSO PRESENT:

BEN HAMLIN, MPH, National Committee for Quality Assurance (NCQA)

TODD LEE, PharmD, PhD, American Board of Medical Specialties (ABMS)*

TOM LYNN, MD, Ingenix

ROBIN WAGNER, RN, MHSA, American Board of Medical Specialties (ABMS)*

KEVIN WEISS, MD, MPH, American Board of Medical Specialties (ABMS)*

* Participating via telephone

C-O-N-T-E-N-T-S

P-R-O-C-E-E-D-I-N-G-S

8:30 a.m.

MS. TURBYVILLE: I wanted to go over quickly some of the parking lots and recurring themes. There is just a few, to make sure I've captured them from yesterday, and then because we're starting with the ABMS measures, we thought it would be a good idea, especially in areas where the specifications are similar across the measures which include the data protocol steps, which is the data cleaning, and others, kind of recap that.

So hopefully then the focus, without ignoring how your evaluation today will influence that, would be on the clinical components and what's different about the measures that you're going to review today, and we had talked about that with Jamie and Jeptha.

So I'm hoping everyone is in agreement with that, kind of built on efficiencies our of what we learned yesterday.

Yes? Okay, great.

So in addition to some of the -- I'm going to start with the parking lot issues. Give me one second, to pull it up. Yes, please?

DR. MARWICK: Could I just ask something?

MS. TURBYVILLE: Yes.

DR. MARWICK: Just in relation to the stroke Ingenix document, it has the same problem that we dealt with yesterday with the acute MI document. The third Ingenix, the chronic coronary disease, I think, is less of an issue.

But I wonder if we should have a discussion with them as to which they would -- if they wish to proceed. I'm not the primary spokesperson on that, but it's just an observation.

MS. TURBYVILLE: We could really -- I mean, I think it's worthwhile for us to examine how to prevent reviewing a measure

that we feel is not going to make it much further through the process. However, the problem that you identified there really is what we rely on the clinicians to do.

So maybe if you guys have a side -- you know, maybe if you and the lead discussant or the lead -- I can't remember who the lead discussant is, feel that it really has these flaws, then, yes, I'll need your input, and at least the lead discussant, in order to then share that with Ingenix. Does that make sense?

CO-CHAIR CURTIS: I think since we're not going to anticipate getting to that today --

MS. TURBYVILLE: We can do it today.

CO-CHAIR CURTIS: -- anyway, maybe offline --

MS. TURBYVILLE: Yes.

CO-CHAIR CURTIS: -- and then we can discuss how to address that --

MS. TURBYVILLE: Yes.

CO-CHAIR CURTIS: -- and get -- pull the measure developer into that discussion.

MS. TURBYVILLE: Yes, okay.

DR. PALESTRANT: I'm the lead discussant on that. I think that there are certain issues, but I guess the question is, are we going to review things that we have -- that may have issues, that may not get full endorsement or pass on those until they come back, but from what I understand, we need to give some guidance of why we're not endorsing them.

MS. TURBYVILLE: We're having a hard time hearing you, David, again. I'm sorry.

DR. PALESTRANT: Okay, can you hear me now?

MS. TURBYVILLE: Yes, that's much better.

DR. PALESTRANT: Okay, I think the

stroke based resource one, there were a lot of interesting components to it, but there were certainly issues. My sense is that we're here to -- we're not just here to endorse. We should be reviewing things that even though we're not endorsing them, I guess because they're in front of us, we should be reviewing them.

CO-CHAIR CURTIS: Okay, so let's put that in the parking lot for right now, but definitely we'll have some discussions and engage both primary reviewers.

DR. PALESTRANT: Just one other thing on the Ingenix thing. We haven't reviewed an Ingenix one in full, and I reviewed two of them and they are -- once again, I think -- requires reviewing one in full because they're all essentially based on the same kind of -- they're all exactly the same, essentially, in terms of their methodology.

MS. TURBYVILLE: Great, thank you,

David.

So some of the parking lot issues are recurring themes. These are not measure specific that we heard yesterday, is a request that the NQF Steering Committee provide additional guidance and statement that the resource use measures that are publicly used should include sound statistical approaches in their estimation and be transparent.

That administrative data lacks -- acknowledging the administrative does lack clinical detail. This type of lacking of clinical detail can affect the risk adjustment, reliability, and, potentially, ultimately, the validity of the measure. No real solution, but it was a recurring theme.

Disparities by socio-economic, race and ethnicity and resource use and literature really does not currently overlap. So we may need to -- what we had been doing towards the end of the meeting is getting not -- rating, a "not applicable" in the disparity

sub-criteria.

The measures submitted are not providing options for data sources in general, so that sub-criteria has also been a "not applicable," so far.

There was a broad question about all measures relying on coding and administrative data. This goes back to one of the earlier points, and that potentially, any source of data that may be influenced by measures, can then also, in some ways, influence their continued validity.

There was also a request, as we move forward, to think about the number of sub-elements that map back to the broad criteria and think about how we might parse it out or reduce that.

So any other over-arching themes or parking lot issues that I may have missed or should be -- now that you've had a chance to think about yesterday, that should be added?

Okay, great. So in thinking about -- do you want -- we can go over what we heard, as kind of the over-arching themes -- do we have the voting of the ABMS measure summary?

Okay, while Sarah pulls it up, staff can summarize what we heard. Overall, as we reviewed the ABMS measures, or Jeptha or Jamie as co-Chairs, if you prefer us to do it, that's fine.

CO-CHAIR CURTIS: You can do it.

MS. TURBYVILLE: Okay, so just give me one minute, and Sarah is going to pull up the -- you know what?

Okay, so, for the ABMS measures, there was -- for all of them, a request for the information on the risk adjustment fitting, how it fit, R-squareds to be submitted.

So while they described some of their risk adjustment approach, there was a request to clarify that in general for the

specifications, including components of the pricing, some time frames that didn't quite always synch up to the amount of data that they requested, that those need to be synched up, for example, in the cardio measure, making sure that they're specifying three years of data because that's what the measure actually requires, and actually when we've reviewed across all of those measures, that was something staff noted, as well, that needed to be more consistent.

I'm just going to -- so, all the issues in the data protocol were similar, and as well as the specifications, really, a need for more clarity on the specifications, as well.

Usability consistently was not applicable across the ABMS measures because they had not provided any information in that section, and feasibility issues -- what was the -- can you scroll down?

So an agreement that because there

aren't administrative data, they are routinely generated. The data elements were available.

There was some concern about susceptibility, the inaccuracies. It wasn't really clear how much they had done to respond to that, and the data collection strategy also had some concerns from the members because it had not been implemented as of yet, and I think it also reflected the need for the specifications to be clarified better, in order for it -- for you to have more comfort in it being implementable.

CO-CHAIR CURTIS: But in general, I think it's fairly consistent across the two measures that we've reviewed, at least with usability, feasibility --

MS. TURBYVILLE: Yes, and feasibility --

CO-CHAIR CURTIS: -- and were there any --

MS. TURBYVILLE: Yes, I think where we saw the biggest difference, if you

could scroll up -- is the specification.

When we went into the -- when we went into the second measure, it was uncovered that there was a need for clarity, and they would have been issues that would have been included, for example, the costing method, in the previous ABMS measure.

It wasn't so much about the components that are actually different from clinical area to clinical area, and importance was also, I think, quite similar, in your findings, yes.

So it was really around uncovering the fact that some of the specifications were not as clear as they should be, and I think that influenced some of the voting, and the need -- yes, because the risk adjustment and the need for the goodness of fit of that was discussed in the first measure.

CO-CHAIR CURTIS: Okay, so I think with that review, maybe we should move on to the zero to 30 days. Mary Ann, if you could

take us through that.

MS. CLARK: Okay, this measure is acute myocardial infarction episode of care for 30 days following onset, and the statement is even a little bit fuzzy, but they go further down and explain what they mean by the time frame. So I'll get into that in a minute.

This is, as we mentioned, by the American Board of Medical Specialities Research and Education Foundation. So the description of the measure is resource use and costs associated with acute myocardial infarction episode during the acute period, and the acute period being defined as, and again, a little bit fuzzy here, but 30 days following initial hospitalization for an AMI event.

An index AMI event identified in all AMI related services are identified in the 30 days following the onset of the acute event. Total AMI related services are

calculated for each patient and summarized at the attributable hospital level, and observed costs are compared to risk adjusted expected costs.

DR. WEINTRAUB: So this includes the initial hospital?

MS. CLARK: Yes, I believe it does. If we go down to the time frame, that sounds like that -- it does include that. So it's not really 30 days post-discharge. It's 30 days from --

DR. WEINTRAUB: The onset?

MS. CLARK: Yes, right.

CO-CHAIR ROSENZWEIG: It's the onset of the admission or the onset of the event?

MS. CLARK: The admission to the hospital, yes.

CO-CHAIR CURTIS: Can the measure developer just confirm that because that's pretty critical?

MS. CLARK: Yes.

DR. WEISS: Yes, it's from the date of admission.

CO-CHAIR CURTIS: Thank you.

MS. CLARK: Okay.

DR. WEINTRAUB: So that would include what happens in --

MS. TURBYVILLE: Microphone.

DR. WEINTRAUB: My problem continues. So that includes what happens in the emergency department, yes?

DR. WEISS: That is correct.

CO-CHAIR ROSENZWEIG: Oh, I'm sorry, and if the patient comes into the -- to see a physician and it's determined that an acute myocardial infarction has occurred at some in-determinant time prior to that out-patient visit, that -- those people are not included in this -- in this particular protocol, I assume, is that correct?

DR. WEISS: Yes, that's correct. This is -- this episode is triggered by an in-patient event.

CO-CHAIR ROSENZWEIG: Okay.

DR. WEINTRAUB: Suppose someone has an MI as a complication of non-cardiac surgery. So they have gallbladder surgery and then have an MI, is that included?

DR. WEISS: If the event ends up as a hospitalization with a diagnostic -- set of diagnostic codes that meet our entry criteria, it would. But if the primary diagnosis is for gallbladder procedure -- or sorry, for something else, and then the myocardial infarction happens, it's possible that that event would not be captured. It depends on the set of codes that are used as part of the hospitalization.

MS. CLARK: Yes, I think it's the principal diagnosis code.

MS. BOSSLEY: Is it just the principal, though?

DR. WEISS: Yes, it has to be the first. I'm sorry, I didn't clarify that. It has to be the first.

DR. WEINTRAUB: So then we are a little subject to the variations in coding, though.

MS. CLARK: Well, yes, I mean, according to the way they're supposed to code it, they're supposed to code it based on the discharge diagnosis, the main thing that occurs during the hospitalization.

CO-CHAIR CURTIS: And I think the instructions of the -- the principal diagnosis for which the patient was admitted to the hospital?

MS. CLARK: No.

CO-CHAIR CURTIS: No?

MS. CLARK: No, that's on discharge.

CO-CHAIR CURTIS: It's determined on discharge, but I think --

MS. CLARK: Not the main reason.

CO-CHAIR CURTIS: Okay.

MS. CLARK: Well, anyway.

CO-CHAIR CURTIS: Anyway, every

measure of MI has the same --

MS. CLARK: Yes.

CO-CHAIR CURTIS: -- that uses administrative data, it uses the same approach. So I don't think we can distinguish on that.

MS. CLARK: Okay, so, let's see, I guess we'll move on down to the data, the area of high impact, I guess. What page is this on? This is on -- oh, it starts on page -- well, I guess I have it on page four.

But anyway, high impact, obviously, this is a high impact area, and they support that with their same literature that was in the other measures, in terms of -- let's see, you know, high impact.

So in terms of opportunity for improvement and the data on that and summary of variation across providers and variation across population groups. You know, again, this is the same information, really, that was presented in the other measures, I believe.

So there was, you know, quite a bit of -- of data on different disparities by population group, and then they had some citations, also, for variation and costs across providers, as well. So it's the same citations.

Let's see, moving on to 1C, which is the measure intent, the purpose and objective of the resource use measure and the components, and the construct for the resource use and costs are clearly described.

Let's see, so, that's on page eight, I believe, and I think that -- let's see, so, the intent is that the measure will be used along with the measure of the 30 day re-admission -- re-admissions to -- to examine the overall efficiency of health care being provided to patients with an AMI.

It will help to identify hospitals that may be undertaking best care practices through identification of those facilities that provide efficient care by examining the

resource use as well as the re-admission rates. So they're saying that these two measures would be put together to really examine the whole AMI episode.

The resource use service categories, I believe they're the same ones, as on the other measure.

Let's see, where are we, now? This is kind of out of order. Let me scroll through the -- well, anyway.

I think the -- what page is -- does anyone know what page that is on because it's not the same?

DR. HWONG: The categories are listed at the bottom.

MS. CLARK: Okay.

MS. TURBYVILLE: So that would be the last sub-criteria for importance.

MS. CLARK: Yes, 1D, right, but it's not in order, on the --

CO-CHAIR CURTIS: It doesn't correspond to any specific elements within the

-- MS. CLARK: Okay.

CO-CHAIR CURTIS: -- application. It's sort of --

MS. CLARK: Yes.

CO-CHAIR CURTIS: -- measure intent, combined with construction logic.

MS. CLARK: Just a general overview of the logic, yes.

Okay, so I think that's about it for resource use evaluation, resource use measure evaluation criteria, the 1A, B, C, and D. Should we --

CO-CHAIR CURTIS: Let me just ask a clarification.

MS. CLARK: Yes.

CO-CHAIR CURTIS: So the measure intent here is different than what we've seen for their other measures, and it's odd because it's -- to me, it specifies that it's paired with the 30 day re-admission measure, and I'm not sure --

MS. CLARK: Yes.

CO-CHAIR CURTIS: -- what that is. To me, that's two different measures of resource use, one of which is embedded in the other.

So I guess they're complementary. I wouldn't see them as being additive, but --

MS. TURBYVILLE: And just as a process clarification, we're not evaluating paired measures. So they would have to be evaluated independently and complementary may be a more appropriate way to frame it.

DR. WEISS: So I can clarify that they are not intended to be paired. It was simply our intent that they could eventually be put together to evaluate efficiencies. This measure is simply intended to focus on the resource use during the 30 day period following an initial hospitalization.

MS. CLARK: Right, and it seems like they would have to be -- almost have to be independent because in the other measure, which is the 31 to 365 days, the -- all of

those events would be related to the first -- the first AMI hospitalization, and whereas in this case, each individual AMI hospitalization is looked at separately.

So there is kind of like, some of the ones that -- some of the ones in this measure could be in that 31 to 365 day period, you know, there would be an overlap, right? That's a question for the developer.

DR. WEISS: I guess I'm confused. Can you just --

MS. CLARK: Sure, so in the other measure, we're looking at patients who have an -- an AMI, and let's say, it's in 2009, and then we're looking from 31 days to 365 days out to look at their costs.

So if they happen to have another AMI episode within that 31 to 365 days, their costs are captured within that first measure, and then independently we're also counting the cost of that -- that second AMI episode, as a separate episode, only a 30 day episode,

right? Is there any issue with that?

CO-CHAIR CURTIS: So you're basically just saying could the same patient enter this measure more than one time?

MS. CLARK: Yes.

CO-CHAIR CURTIS: And then cross-cut it in a separate measure, of 31 to 365?

I think potentially, depending on how it's specified, unless there is a --

MS. CLARK: Yes.

CO-CHAIR CURTIS: I guess we get into that in the data specifications, if there is an exclusion for one per calendar year.

MS. CLARK: But it doesn't seem to be an issue, probably.

CO-CHAIR CURTIS: Okay.

MS. CLARK: So any other questions? Well, are we ready to vote then?

1A, which is the measure focus addresses a specific national health goal priority identified by DHHS or the National Priorities Partnership convened by NQF or a

demonstrated high impact aspect of healthcare.

Okay, 1B is demonstration of resource use or cost problems and opportunity for improvement. Data demonstrating variation and the delivery of care across providers and/or population groups. So, again, these were the same citations as in the other measure.

DR. HWONG: I think the comment from the previous measure was that a lot of the citations were not specifically about that 31 to 365 day period. Were the citations more relevant to this 30 day period?

MS. CLARK: No, I think they were the same.

DR. HWONG: Okay, that's fine, then.

DR. WEINTRAUB: I think the --

MS. TURBYVILLE: Microphone.

DR. WEINTRAUB: -- the evidence --

MS. TURBYVILLE: Microphone.

DR. WEINTRAUB: I think the

evidence for variation here is probably less. I mean, we know that when people go home all kinds of things happen that are variable.

When people are treated for MIs today, the course of treatment is so firmly within guidelines that to step out of that is a little more unusual. I think that there's probably less concern here than there would be in some of the other measures.

Now, that doesn't mean there is zero, but I'm not sure that they've demonstrated -- that they have demonstrated that this is -- this is a critical national need to look at resource variation and treatment for acute MI.

MS. CLARK: Okay, shall we vote, then?

CO-CHAIR ROSENZWEIG: Well, there were some, actually, some references that did --

MS. TURBYVILLE: Microphone.

CO-CHAIR ROSENZWEIG: I'm sorry.

They did cite some references that referred to variation in care for the MI in the hospital, which actually is within that 30 day period.

They talk about -- there is some references related to utilization of coronary angiography, variations and -- and institutional variations in length of stay, and complications of -- of MI, as well.

So they do address some of that, at least, if you consider that to be -- you know, it is -- since the actual clinical in-patient event is actually a part of this area of study. So they're a little more specific to that than in the other protocols, I would believe.

CO-CHAIR CURTIS: So we have to re-vote on that?

MS. TURBYVILLE: We're going to re-vote on that.

MS. CLARK: The timer got started before you -- so we're going to re-vote?

MS. TURBYVILLE: We're going to

re-vote on 1B.

MS. CLARK: Okay, 1B re-vote.

MS. TURBYVILLE: Okay.

CO-CHAIR CURTIS: So moving to 1C?

MS. CLARK: 1C, the purpose objective of the resource use measure, including its components and the construct for resource use and costs are clearly described.

CO-CHAIR CURTIS: The only feedback I would personally give to the measure developers, I think that in the measure intent, describing it as being paired with the re-admission measure is what lowered, at least, my vote on this particular one.

MS. CLARK: Finally then, 1D, which is the resource use service categories that are included in the resource use measure are consistent with the -- and representative of the conceptual construct represented by the measure.

So, again, these were the same resource use categories as before.

CO-CHAIR CURTIS: And just how do we vote on this -- for the other two measures? I think it's probably the same.

DR. HWONG: I think it was --

CO-CHAIR CURTIS: Sorry, did we complete the vote?

MS. WILBON: I started it and then --

CO-CHAIR CURTIS: Okay.

MS. WILBON: -- you started talking so I restarted it. Sorry.

CO-CHAIR CURTIS: Ashlie is trying to move it along. Appreciate that.

Okay, so moving on to scientific acceptability.

MS. CLARK: Okay, the long one.

CO-CHAIR CURTIS: Walk us through.

MS. CLARK: Okay, so S2 -- sub-category S2, I guess, general approach.

So when I was reading this and the other ones, as well, I was maybe -- maybe people have approached this differently, but

the general approach that they describe is really more of a process for how the measure was created, and not a general approach to the method. So I don't know if that's really what was supposed to be put in this group -- in this -- as an answer to this question, or not.

But they talk about, you know, consensus panels and clinical input and all of that, as the way that they got to develop the measure. But I was, I guess, assuming that it was going to be more of a general description of how the measure worked. So I don't know which one -- is it -- does it matter?

MS. TURBYVILLE: Yes, I would focus more on the other data elements.

MS. CLARK: Yes.

MS. TURBYVILLE: We can work with them to --

MS. CLARK: Okay.

MS. TURBYVILLE: -- update that document, but it won't necessarily --

MS. CLARK: Yes.

MS. TURBYVILLE: -- affect the specs of the submitted --

MS. CLARK: Okay, so the general approach --

MS. TURBYVILLE: Good point, though.

MS. CLARK: -- you know, again, is basically exactly the same as the other one.

Type of resource use measure. So, again, they just say per episode, which I guess I would like a little more description, there. So it's really, what is an episode because an episode can be anything.

So I would recommend a little bit more description on that, that it's within that certain time period, the initial hospitalization through 30 days.

MS. WILBON: Mary, just a quick note. That -- there is a few options that we give them in the form, on the electronic submission form, that are just check boxes that help feed our database so we could kind

of search for them at a later time. That's one of those fields.

So just so we can kind of have an identifier for the measure.

MS. CLARK: What are the other choices?

MS. TURBYVILLE: So S41, S42, S3 - -

MS. CLARK: No, no, the other choices --

MS. WILBON: No, the other choices --

MS. CLARK: -- per capita --

MS. TURBYVILLE: What are the other choices of S3?

MS. WILBON: Type of measure, type of resource use measurement.

MS. TURBYVILLE: Per capita, per episode, procedure, so it's different types of resource use measures that they might be focusing on. These are episode based measures.

MS. CLARK: Target population, they left blank, but I guess we'll get to that.

CO-CHAIR ROSENZWEIG: It seems like it's left blank in most of them.

MS. CLARK: Yes.

MS. TURBYVILLE: Yes, again, it's a standard list, and some of them just don't touch on --

MS. CLARK: Okay.

MS. TURBYVILLE: -- and it's a list that goes across all measures, that's for kind of out-facing NQF tool, and these measures don't necessarily just focus on one of them listed.

So I would -- that is --

MS. CLARK: Okay.

MS. TURBYVILLE: -- intentionally left blank, by some.

MS. CLARK: Okay, data dictionary or code table, I mean, I think this was pretty much the same as the other ones, as well. I

mean, I kind of was wishing there would be a little bit more description in that table in terms of definitions of the variables. I mean, it was pretty generic. So that was my only comment there.

Let's see, data protocol, so preparing the data for analysis. So this is, again, on the data cleaning, and they're suggesting a guideline, as opposed to a, I guess, specification. So standard approach to cleaning the claims data that payers are using today, I guess.

Let's see, I don't think they're really recommending any -- let's see, if organizations impute missing data, they're saying to not use imputed data. So that would be one recommendation that they make.

DR. WEINTRAUB: And that's the same as the --

MS. CLARK: Yes.

CO-CHAIR CURTIS: I think just highlighting the different --

MS. CLARK: Okay.

CO-CHAIR CURTIS: You know, the differences would be --

MS. CLARK: Yes, so it's the same -- so do we need to -- well, when we get to the ratings, we can go through those.

Okay, data inclusion criteria, so it's the same type of thing. I think it's exactly the same, right? Paid claims with non-missing enrollee identification numbers, blah, blah, blah.

So data exclusion criteria, those are the same, as well.

DR. WEINTRAUB: What page are you on?

MS. CLARK: This is on page 11, S63. Missing data, I believe that's the same, as well, and then the data type, and the administrative claims, as we know, and then they also have "other" because of these pricing files that they're using, I believe.

Data source or collection

instrument, we already talked about. Data source, okay, so, now, we're getting into the clinical framework.

Okay, the brief description is that resource use and cost associated with the AMI during the acute episode, and, again, defining that as the 30 days following initial hospitalization.

So, again, I think maybe if they want to be a little bit more specific here, that it includes the initial -- the index hospitalization, as well.

So it's really 30 days from admission for the AMI, I believe, right?

CO-CHAIR ROSENZWEIG: It looks to me like they're saying -- maybe I'm wrong, but the way I read it, is that it's 30 days following -- well, oh, I'm sorry, they say hospitalization for an MI event.

MS. CLARK: Yes.

CO-CHAIR ROSENZWEIG: Okay, I'm sorry. But then they say the event is

identified in all AMI related services that are identified in the 30 days following --

MS. TURBYVILLE: You need to speak up.

CO-CHAIR ROSENZWEIG: What?

MS. TURBYVILLE: Use the microphone.

CO-CHAIR ROSENZWEIG: What?

MS. TURBYVILLE: Your microphone is back on. You're good.

CO-CHAIR ROSENZWEIG: Okay, all right. It's just that there is a little confusion here.

MS. CLARK: Yes.

CO-CHAIR ROSENZWEIG: In one case, they're saying 30 days following the hospitalization, but then in another case they're talking about 30 days following the onset of the acute event, and, as we know, the onset of the acute event might occur a few days prior to the hospitalization, or it might occur after the hospitalization, right? I

mean, the --

CO-CHAIR CURTIS: Right, so but I think the --

CO-CHAIR ROSENZWEIG: So the 30 --

CO-CHAIR CURTIS: The clarification that they gave us was that the triggering event was the date of admission. So I assume that their calculations start at the -- from the emergency department through that initial hospitalization.

CO-CHAIR ROSENZWEIG: All right, well, then they need to change the --

CO-CHAIR CURTIS: Right.

CO-CHAIR ROSENZWEIG: -- the way they write it here.

CO-CHAIR CURTIS: Yes.

CO-CHAIR ROSENZWEIG: Okay, it's not that -- if they clarified it as such, that's fine. It's just that it's not written as such -- CO-CHAIR CURTIS: Yes.

CO-CHAIR ROSENZWEIG: -- in the specifications of the measure.

DR. WEINTRAUB: That's correct.

CO-CHAIR ROSENZWEIG: And that has to be clarified.

MS. CLARK: Yes.

DR. HWONG: Mary Ann, I had -- and also for the measure developer, some questions in terms of this clinical framework, in terms of the eligibility criteria, and I'm wondering if there are some ways to make this a little bit clearer because I think they mentioned that eligibility, you know, has to be for the previous year and the current year.

Like, there is -- I think there is a statement somewhere that, you know, you need to be eligible, yes, in the prior year and the current year. And so the question I have is if you're looking at an event that's happening in the measurement year, and you're only looking at 30 days following that, why would you need to have the eligibility criteria be a full year?

I mean, it's just sort of a

question. You'll -- in so doing, you end up decreasing your sensitivity. You'll lose, you know, potential cases that you'd want to include because that eligibility criteria for the measurement year becomes too stringent.

So I'm just -- you know, maybe for -- I don't know if you had some perspective on that, Mary Ann, or the measure developer could answer that. But you potentially are losing cases that you could count.

MS. CLARK: Right, yes, I think we're -- that's a good comment. Does the measure developer want to comment on the reason for requiring a full year worth of data?

DR. HWONG: Or eligibility, yes.

DR. WEISS: Yes, sure. This is actually a topic that we debated quite a bit within our development group, for not only this measure, but for some other measures that focus on a non-365 day period.

The decision ultimately came down

to try to be consistent across all of our measures, to make it a bit easier on folks that would be implementing these measures, in that if you're assessing eligibility criteria, you can do it -- across all of the ABMS-REF measures.

And we realize that we're going to lose sample size and exclude cases because of this eligibility criteria, but it was more a decision of pragmatism than anything else.

DR. HWONG: Got you. You know, my only feeling on that, you know, again, coming from, you know, health plan and understanding sort of enrollment, and having someone with two years of continuous enrollment, actually, you know, is not an easy thing.

You will actually lose, you know, a large number, and especially if the measure period of interest is only 30 days, you really have a chance to gather a lot more cases.

So, you know, if the measure developer could contemplate that and think

about maybe, you know, reducing the stringency of that criteria, I think you could actually apply this to a lot more individuals.

So the other thing, one other thing in terms of, again, this is sort of the eligibility and sort of measure construction, but you know, if the measure event, you know, the triggering event is supposed to occur between January 1 and December 31st, it can't really be December 30th because you still need that 30 days follow up to actually assess, you know, the resource use during that period of time.

So I would ask maybe the measure developer to clarify that, that you would have to set that, you know, to like, you know, December 1st would be the final date that you could actually submit a triggering event.

Is that how you -- you know, maybe that would --

DR. WEISS: Yes, we can make that clarification. You're absolutely right, you

have to have full capture of the 30 day follow up period, within the dates.

DR. HWONG: Okay.

MS. CLARK: Yes, usually, when I -- I've done a lot of these claims and analysis studies, I always do it from the time period from the -- you know, each patient has an index date.

So you're doing it from their own index date. So, I mean, that's another approach, as well.

CO-CHAIR ROSENZWEIG: Just with respect -- I have a question, maybe other people know more about this than I. But now with the new healthcare law that specifies that you don't -- you know, that you can't be prohibited from joining a plan due to a pre-existing condition, so will there be a lot of patients who actually join a plan on the -- you know, when -- with the onset of an acute MI, and would there be a bias related to individual -- you know, which plans people

decide to join, or is that -- this is not an issue, here?

CO-CHAIR CURTIS: I think the requirement that you have at least a year before probably makes that not relevant to this, and that you'll have equal amount of time for risk adjustment. I don't know how often people do that or would want to do that, but I don't think it's necessarily relevant.

CO-CHAIR ROSENZWEIG: So that is a good reason for having the patient in the plan for a year, prior to?

CO-CHAIR CURTIS: Right, you have to have a stable, you know, time of -- for which you can obtain the information --

CO-CHAIR ROSENZWEIG: Okay.

CO-CHAIR CURTIS: -- about co-morbidities and cardiac status.

CO-CHAIR ROSENZWEIG: So that's a good rationale for that, okay.

MS. CLARK: Let's see, so going on then in terms of the in -- the clinical frame

work. So, basically, the age range is the same as the others, 18 to 85, during the measurement year.

DR. WEINTRAUB: In the re-vascularization one, there was no upper age. In the MI one, there was. So you have some variability.

MS. CLARK: Right, and I think our comments on the previous AMI measure was that, why? Why have the upper end, and it was because, I believe, the response was because the costs could be higher for elderly.

CO-CHAIR CURTIS: The sample -- process might be different.

MS. CLARK: Yes.

CO-CHAIR CURTIS: With a clinical decision.

DR. WEINTRAUB: Well, the sampling -- it doesn't entirely makes sense to me. MI and revascularization, they're in -- pretty much in the same age range, and you have plenty that are in the upper age range.

The reason for not including them in the super elderly was because this was -- this was aimed at private plans, rather than Medicare. So the issue, to me, is the same, here.

MS. CLARK: So in this --

DR. WEINTRAUB: Hardly the most important thing in the whole world, but --

MS. CLARK: In this commercial database, though, they're including Medicare Advantage. So is that -- was that excluded? Well, it was, according to this criteria, I guess.

I mean, is this measure intended to be applied just to a commercial under -- non-Medicare population, or is it including the Medicare Advantage population, which is commercial?

DR. WEISS: In testing, it was primarily in a non-Medicare commercially insured population. There were very few people that were over the age of 65.

However, we did test this and the post acute AMI measure in a sample of Medicare data, so it's not the intent that it should only be applied to commercially insured populations.

MS. CLARK: No, I think in this measure, when they do the testing, they also tested it in Medicare claims.

DR. MARWICK: Could we just clarify that the emergency room costs are incorporated in this?

People admitted to hospital, after their emergency room stay because there may be -- for example, with use of CT, there may be significant costs there that might not be captured.

MS. CLARK: Normally, if they're admitted after an ER visit, that's included in the hospitalization.

DR. WEINTRAUB: That is what they said in response to -- I think one of the places you could have problem is if someone

dies in the emergency department. They're not actually admitted.

So with -- when people who actually don't get as far as being admitted, are they included?

DR. WEISS: No, you have to be discharged alive.

DR. WEINTRAUB: Oh, that's interesting. Why that choice because that's going to take out for approximately five percent who die?

DR. WEISS: Those people don't have costs in the 30 day follow-up period, and I understand that they might have very high hospital costs, but we're trying to look at a population of patients that might consume resources over that post-follow-up period.

DR. WEINTRAUB: Yes but, you know, that creates quite a bias.

DR. WEISS: It's the same bias, though. There is no differential bias. Everyone is discharged alive.

So, we understand that we may be underestimating overall resource use, but this isn't a measure intended to say, "Hey, how much do in-patient events totally cost?" We're trying to figure out if there are differential resource uses across entities, and in this case, hospitals.

DR. WEINTRAUB: All right, let me explain why there might be a bias. If you have hospitals that aren't that -- that aren't doing a good job, and their sick people dying, they're taken out, those are the people that might use a lot of resources.

DR. WEISS: Yes, I agree -- I acknowledge that fully.

CO-CHAIR CURTIS: And I'll just to ask the comparing question, if someone is in the hospital for 30 days, thus not discharged, they would be excluded?

DR. WEISS: I'm sorry, I couldn't hear that question.

CO-CHAIR CURTIS: All right, in

the case where a patient is in the hospital for 30 days post-MI, rare, but it does happen, they would be excluded?

DR. WEISS: Yes, you know what, our measure specification does not do a good job of dealing with that instance. So, the way that we've written it, it would be excluded -- well, you know what? I would have to -- we'll have to clarify that.

It wasn't the intent that those individuals would be excluded, but I could see in operationalizing our specifications, how there could be variability around that case, because if that person is discharged alive, we would want to capture their costs.

CO-CHAIR ROSENZWEIG: What if they die in the 30 day period, even if they're discharged alive?

DR. WEISS: They're included.

CO-CHAIR ROSENZWEIG: They're included.

MS. CLARK: Any other questions,

there?

Okay, let's see, so, again, they're excluding hospitalizations that were subsequent hospitalizations, so the diagnosis code with the 410-X2.

Let's see, there are other -- same exclusions as in the other measure, in terms of patients with active cancer, end-stage renal disease and let's see, what else? Some of the other organ transplants, HIV.

Also, discharges to a skilled nursing facility, excluded, which in this case, I'm wondering if that really makes sense.

CO-CHAIR CURTIS: I think this is the reason why it was in the other one, is that they thought that they couldn't adequately or accurately capture the resources used in that setting.

And so, it's a challenge, but I do -- we questioned it on the last one, I think, I personally would question it, on this one,

but it's no different.

But given the variation in discharge across the nation, I kind of wonder if we're attributing this to the hospital level, if that's appropriate, and I think ultimately this is attributed to the hospital level, not to the physician.

MS. CLARK: Yes.

DR. MARWICK: The impact of that is much greater with this, obviously, isn't it?

CO-CHAIR ROSENZWEIG: Would hospice care be an exclusion?

DR. WEISS: No.

MS. CLARK: Can you just remind us of why -- you know, why do you think skilled nursing is going to be difficult, as opposed to hospice, for example? I mean, what's the difference there?

DR. WEISS: Well, hospice, you can still, in most data systems, observe the care that's being provided, if they're seeing physicians, if they have an area home nursing

agency coming in. Those are claims that are submitted and observable.

If individuals are admitted to a SNF, often times that may result in them moving outside of the data stream that we're able to capture, especially in our test data set, and then they have this large, immeasurable period.

And so, with that immeasurable period, we're potentially, again, introducing another bias, and so our approach was then to exclude that in-measurable period.

I mean, this is the same concerns around the sicker patients that die. Maybe sicker patients that were admitted to a SNF, and if those are the higher resource patients, then they're a potential directional bias.

MS. CLARK: So, the market --

DR. WEISS: So we wanted to try to avoid that.

MS. CLARK: The market scan data doesn't have skilled nursing claims. That's

basically what you're saying.

DR. WEISS: That is correct.

MS. CLARK: So, the --

DR. WEISS: But similarly, if you were using Medicare data and they were put into -- admitted to a skilled nursing facility, and then Medicaid became the primary payer, it may be difficult, measuring that, as well, unless you had a combined data set.

This was more of, again, an ability to measure costs during that period than anything else.

MS. CLARK: Right. Okay. Let's see, in terms of again, identifying the event, it's an AMI diagnosis code on admission, principal diagnosis code, right?

The events within the 30 days would be anything related. Again, that's the same, I think, definition of AMI related codes.

So, it's going to be the same diagnosis codes and DRGs, if they were

admitted to the hospital, including anything for unstable angina, arrhythmia, pace makers, heart failure, atherosclerosis and procedures that would be related, as well.

And so we had some comments on the coding that needed to be updated here. So, that would need to apply to this, as well.

Pharmacy related -- AMI related medications, I think those were the same ones, as well, right, and we had some comments on those, before? Yes.

Okay, so, let's see, anything new here? First event, includes any 30 day period as a triggering event for the episode.

Again, I think this needs to be further clarified or defined here.

Length of stay --

DR. WEINTRAUB: What page are you on?

MS. CLARK: I'm on page 14. Length of stay; for an event to qualify for initiating episode, the length of stay needs

to be more than one day.

Is that something that is --

DR. WEINTRAUB: Well, if they're going to stick with the idea that they have to be people who are discharged alive --

CO-CHAIR CURTIS: Microphone.

DR. WEINTRAUB: If they want to include people who are -- only people who are discharged alive, it's reasonable.

If they -- if we -- if they change it to include people who die, then this doesn't work.

MS. CLARK: Right. Okay. Yes, and then the next paragraph is the discharge alive. It only includes people that were discharged alive.

CO-CHAIR CURTIS: I guess, I'm just having -- and we discussed it. I just want to touch on it one more time.

But it doesn't necessarily make sense to me that, specifically, also given variation and length of stays across hospitals

and aggressiveness of pushing people out, that if someone is in the hospital for two days and gets discharged on day four, or sorry, gets discharged and dies on day four, how is that different conceptually, than someone who is in the hospital and dies at day four?

And you could have -- so, I think the problem of deaths doesn't stop at discharge, and if you're going to include them, once you are discharged, I think you probably have to consider including them during the admission, recognizing -- and I understand that -- your rationale on that you don't want to reward high mortality hospitals for looking really good on not using a lot of resources.

But on the other hand, that's why ultimately, once these are moved towards value, you would look -- you know, it would hopefully be offset in that regard.

DR. WEISS: Yes, so, let me respond to that, if I can, because I agree

with you, completely.

One of the complexities of measuring this and identifying that death on day four is that a lot of especially commercial claims data, do not do a good job of identifying mortalities.

It does do a good job of identifying mortality in hospitals. So, there is ways to say, "Hey, look, this person died in the hospital," through discharge codes.

We cannot reliably identify mortality that happened outside of the hospital, in a lot of these commercial claims. That's an exception, if you move to a database that does have mortality information, and perhaps, that's something that we should reconsider, in light of those type of databases.

But again, there was a balance here, in what we can measure and still try and be consistent within the measure.

DR. WEINTRAUB: That's true, but

I'm not sure that's a good reason for -- that particular issue is not a good reason for excluding the deaths in the hospital. That would be --

DR. WEISS: No, no, I'm sorry --

DR. WEINTRAUB: That would be an argument for including them, seeing is you have other people that are dying.

DR. WEISS: I was just speaking to the rationale for not excluding somebody that died day four, post-discharge.

We can't measure it in the data that we tested our measure in.

CO-CHAIR CURTIS: The other thing I noticed on the exclusion criteria, that sort of varies across the different measures, is the exclusion or inclusion of pregnant patients.

So, on the repost revascularization measure, it looked like you were trying to exclude pregnant patients, and in the MI measures, both of them now, you are

including that population.

Obviously, it's a low frequency event, but just didn't quite understand the difference in decision making.

DR. WEISS: So, that's driven by separate clinical work groups, across these two measures.

Our AMI measures were one clinical work group. Our CAD measures were a separate clinical work group, and they made separate clinical decisions and we did not try to reconcile some of those clinical decisions across measures.

DR. WEINTRAUB: So, I mean, there is a clinical scenario in which this occurs, in young women, and that's coronary dissection in the peripartum period.

They can present with MIs, and they both could be revascularized.

CO-CHAIR CURTIS: We'll just accept that at, you know, decision and we'll take each measure separately. Thank you for

the clarification, though.

CO-CHAIR ROSENZWEIG: Your rationale for excluding patients with end stage renal disease? I mean, that seems to be a condition in which you have a high incidence of acute MI.

I know there are much higher cost patients, in general, but shouldn't -- is there a clear cut reason for excluding them?

DR. WEISS: Again, this is one of our standard exclusion across all of our measures, because of concerns about differential resource use.

We really were following along with what NCQA does, as part of their HEDIS measures, and their relative resource use measures, because of the concerns around differential costs.

DR. WEINTRAUB: How are you handling transfers?

DR. WEISS: That topic is later on in the specification. I don't know if you

want to talk about it now.

DR. WEINTRAUB: All right.

MS. CLARK: So, in terms of co-morbidities, they're basically handling that through the risk adjustment, the HCC risk adjustment, except they're separating out the heart failure patients.

So, that's the same as in the other measure.

Let's see. Severity -- no severity adjustments. Let's see, no -- yes, now, I'm going onto 17.

Concurrency of clinical events. There is nothing provided there. Measure construction, logic. So, again, you know, the brief overview of the construction logic, identifying the population, related resources, assigning standard prices and creating the episode strata.

CO-CHAIR ROSENZWEIG: Just going back to page 14, with respect to pharmacy, it seems like you have a fairly limited number of

medications that are listed. You're not including anti-arrhythmics?

CO-CHAIR CURTIS: That is consistent across the two measures.

I think we asked them about that on the first 31 to 365, and the response was that they really wanted to focus on the ones that were most likely AMI-related, for something like amiodarone, for instance.

It's a little hard to say, is that AMI-related or afib-related and -- it's complex. So, I think they opted to try and take the most focused list possible, within -- you know, but there are limitations to that decision.

DR. WEINTRAUB: Yes, I think here, it's more problematical.

Atrial fibrillation can occur, as a complication of acute myocardial infarction, and certainly, on a v-tach.

You're going to be giving anti-arrhythmics that are -- that I think are

clearly related. I think excluding them here is a bigger problem.

CO-CHAIR CURTIS: I think it's the same issue. I don't know. I don't think it's different than what we voted on before, or our discussion before.

MS. CLARK: Okay, then moving on, we're on page 18.

So, this is where they do talk about -- let's see, it's identifying patients that are transferred between two in-patient facilities.

Information is used when reporting the results as findings or stratified, by those that were and were not transferred.

So, I guess that's explained in the stratification. We're not to that, yet.

DR. WEINTRAUB: So, I'm not sure what you do with that, though.

Is the -- and how about the receiving facility? Is the receiving facility not included for the MI, at all, and

what happens to -- so, what happens? How do you handle that, analytically?

DR. WEISS: How do we handle what, analytically?

DR. WEINTRAUB: So, someone has an MI, they come to a hospital that doesn't perform revascularization. They're transferred to a hospital that does. How do you attribute that MI? Whether it's stratified or not stratified, how do you attribute it?

DR. WEISS: Yes, so, our attribution logic focuses on the hospital with the majority of the length of stay.

So, if that person stayed for six days and was -- five of them were in the receiving hospital, the attribution would be to the receiving hospital.

MS. CLARK: Page 23, it has this method.

So, if someone gets transferred, then the cost of the initial hospital are just

included -- if the second hospital had the longer length of stay than the initial hospital, those costs would be assigned to the second hospital?

DR. WEISS: That is right.

MS. CLARK: Okay.

DR. WEINTRAUB: So, I mean, there is no perfect way of doing this. But you can see what happens, how you can have a problem. Someone is admitted to a community hospital that doesn't have revascularization, they're there for four days, especially if it's a non-STEMI.

They're transferred -- they transfer, they have the STEMI at the receiving -- they have the PCI at the receiving hospital, and go home the next day.

But it's all attributed to the -- the community hospital. Maybe that's okay, but you know, it becomes a little peculiar.

MS. CLARK: So, in terms of identifying that initial event, if they both

had a principle diagnosis of AMI, are you looking at the same patient, and then -- and looking at like the admission source and the discharge status, because you would need to look at -- to try to determine which one was -- or is that necessary? Maybe you don't need to do that.

DR. WEISS: We look at the discharge status and the fact that two events might be -- there might be a discharge date and an admission date that are exactly the same.

So, we've identified transfer status and then the fact that there is an admission and discharge date that are common, we identify that individual as having been transferred.

MS. CLARK: Okay, all right. Let's go back up to -- so, this is, just again, talking about the specification of the logic. We already talked a bit about this, discharged alive, transfers, eligibility and

continuous enrollment.

So, this is where you had the comment about whether there really needs to be a full year post-AMI. So, just a comment, I guess.

Let's see, anything else that stands out here? I don't think so. Those were -- exclusion criteria, again, I think were fairly similar.

Related resources, we already talked about that, but in-patient hospitalization events, out-patient events, procedures and lab, you know, all the costs associated within that 30 period that would be related to AMI, according to the codes that they identified, and need to be updated.

Measure trigger and end mechanisms, we already discussed that, but some clarity needs to be put around that.

Redundancy and overlap, that is not applicable here, I guess. Complementary services is not specified here, either.

Then we have resource use service categories, those are the same ones. Inpatient facility, evaluation and management, so on.

So, okay, so, emergency department, I think is added here, right? I don't know that -- was that in the previous one, ambulatory services?

Can the developer comment on that? I can't remember whether, on the other AMI measure -- I'm assuming it was, but was emergency department services a specified resource use category?

CO-CHAIR CURTIS: I believe it was. We can confirm. But I think it was consistent.

MS. CLARK: Okay, I just didn't remember seeing that one.

So, let's see, in terms of identifying the categories, they're doing this based on the, you know, codes on the claims, once again.

So, I still have a little bit of a -- an issue with this, because I don't think it's very specific.

I mean, you talk about BETOS categories, which apply to the HCPCS codes and the -- I don't know, it's just not quite -- very clear on how the assignments are being made to the various resource groups, that's all.

I would like a little more clarification, unless everybody -- it's perfectly clear for everybody else.

Care setting, so, here we have, again, ambulatory care, which includes ASC, urgent care, clinician office.

So, I guess I would ask -- these are probably standard categories, is that right, the care settings?

MS. WILBON: Yes, those are standard.

MS. CLARK: So, is there one for out-patient hospital?

MS. WILBON: I don't believe so.

MS. CLARK: And that is not on here, just acute -- it just says hospital acute care facility, so, I guess, in this grouping, is everything done at a hospital just considered hospital? Because they have different settings.

MS. TURBYVILLE: I'll find the list and clarify. Heidi, do you have any, for the taxonomy or care setting, it's out-patient? I'm sure out-patient is on there, right?

MS. BOSSLEY: Yes, it's ambulatory.

MS. TURBYVILLE: It's ambulatory?

MS. BOSSLEY: Yes.

MS. TURBYVILLE: All right.

MS. BOSSLEY: Then there is three sub-settings underneath it, ambulatory, surgical center --

MS. TURBYVILLE: Okay.

MS. BOSSLEY: -- clinician office

and something else. I'm blanking on the third. I'll look it up. I've got it.

MS. CLARK: So, the reason I'm asking is there is a definite, you know, for Medicare, anyway, there is a whole different payment system for hospital outpatient versus ambulatory surgery, free standing ambulatory surgery.

So, you know, the costs are different for those, right?

MS. WILBON: Okay, we'll check on that list and let you know.

MS. BOSSLEY: We didn't distinguish between the two. We had many discussions on the best way to do it, and right now, we don't distinguish between the two, if I remember correctly.

MS. CLARK: Because the cost structures are really different, between the two.

MS. BOSSLEY: Yes.

MS. CLARK: Yes.

MS. BOSSLEY: It was one that we went back and forth on, and probably, I suspect we'll be updating the taxonomy again, to --

MS. CLARK: Okay.

MS. BOSSLEY: -- add it back in, yes. It's not there, now.

MS. CLARK: Okay, let's see, so then moving onto S10, I think we already talked a little bit about that, risk adjustment method, S10.1.

So, this is the same risk adjustment method as in the other measures.

So, it's -- I guess it's again, using -- starting off with using the Medicare HCC method, but then doing some adjustments for -- that are specific to AMI, I guess, and then several different models were tested, and I think we already provided comments on getting better clarity around those models in the R-squareds.

Let's see --

DR. WEINTRAUB: I think they actually lack calibration, as well, but I think that is true of all of their --

CO-CHAIR CURTIS: Right, so, I think we can just sort of have this similar feedback across this.

MS. CLARK: Yes, okay, and then onto page 23, down to 10.2, the stratification method.

Here again, we talked about this, but the CHF group and then the transfers to other hospitals.

So, again, I guess, is there any discussion on why does CHF need to be called out separately? I mean, there could be other groups. Is that something --

CO-CHAIR CURTIS: Right, I think that's the same --

MS. CLARK: Same comments?

CO-CHAIR CURTIS: -- comment we had, is why heart failure as opposed to any one of other comorbidities, I think it's

consistent, without empiric evidence, that this is --

MS. CLARK: Yes.

CO-CHAIR CURTIS: -- the absolute one that had to be adjusted for, it seems somewhat arbitrary.

MS. CLARK: Yes, okay, and then the costing method, I think they are the same comments that we've had on others. They're using the same methodology.

CO-CHAIR ROSENZWEIG: When we talk about stratification method, does it always assume that there has to be adjustment for it? I mean, or in certain cases are they just stratifying to look at different categories?

CO-CHAIR CURTIS: I think there are different reasons for stratification, as we've sort of discussed.

You know, it might be something that helps you drill down on the data. It might be something that you don't think risk adjustment alone can account for. It might be

related to a disparity in care that you don't want to obscure.

But the rationale isn't really provided here, as to why heart failure, as opposed to any --

DR. WEINTRAUB: I mean, theoretically, from a mathematical point of view, the reason for stratification is you believe that there is going to be an interaction with other covariates.

So, for instance, if you believe that the effect of age is greater in patients with heart failure than without heart failure, and you don't want to build a model with interaction terms, since they're always very confusing, then that's the reason for stratifying and doing it, from a mathematical -- doing an analysis.

I don't think they've gotten into that, but in modeling that I've done, these kinds of conversations are very intense, go on for months, trying to figure out what you're

going to do.

CO-CHAIR ROSENZWEIG: Okay, well, earlier on they mentioned that they were going to consider NS STEMIs and N STEMIs separately, so, would that be -- should that be included in this section, as well?

CO-CHAIR CURTIS: Well, they said that they can't, because they can't distinguish between the two, in the data that they have.

DR. MARWICK: But that's even more of problem --

CO-CHAIR CURTIS: It's a very big problem --

DR. MARWICK: -- in this group, than it is in the later group.

DR. WEINTRAUB: Because there, you really do, you may very well have interactions, but you have other things going on, because of cost of care is so very different for STEMIs, and it -- and in the N STEMIs, and if they can't distinguish it, it

is a limitation of this whole process, there about.

CO-CHAIR CURTIS: But again, one that's consistent across all the outcomes measures for MI.

DR. WEINTRAUB: So, you know, I think that they ought to try and justify what they're trying to accomplish a little bit better, within that stratification. I wouldn't particularly stratify heart failure.

The other thing is, heart failure can be a complication, and they make it clear, it's heart failure on admission, or heart failure when -- prior heart failure, if it's heart failure as a complication, it's the same kind of problem they had with stratification that we saw before, the revascularization model, where they were stratifying on events downstream, which makes no sense.

CO-CHAIR CURTIS: So, I believe it was specified, as in the previous 12 months, but could the developer clarify that?

DR. WEISS: It's heart failure identified in the period prior to the index event.

DR. WEINTRAUB: If you're going to do it.

CO-CHAIR CURTIS: Much better.

MS. CLARK: Okay, then moving onto the -- let's see, attribution approach, which is page 28.

CO-CHAIR CURTIS: So, for the costing method, we'll have, again, the same comments that we had before.

MS. CLARK: Yes, and so, attribution method is at the hospital level, as we discussed, with the hospital with the majority of the length of stay during the index AMI, having the -- getting it attributed there.

Peer groups, they don't specify guidelines, or have the guidelines for identifying or defining peer groups. So, nothing is defined there.

So, there is a comment there, though, that says we do not think it's feasible for most users to link with databases that contain hospital information, such as number of beds, teaching status or other criteria. That seems pretty reasonable to me. I do that all the time.

CO-CHAIR CURTIS: Would the measure be helped, in terms of the interpretability or usability of it, down the road, if you did have a peer group?

I, personally, would think that if you had CABG capable hospitals as a peer group --

MS. CLARK: Right.

CO-CHAIR CURTIS: -- that would make sense, or primary for --

MS. CLARK: Yes, right.

DR. WEINTRAUB: Absolutely.

MS. CLARK: I would think so.

DR. WEINTRAUB: Hospitals that are doing revascularization --

MS. TURBYVILLE: Microphone.

DR. WEINTRAUB: Hospitals that are doing revascularization and hospitals that are not are going to have very different cost structures.

CO-CHAIR CURTIS: Yes.

DR. WEINTRAUB: Necessarily so.

MS. CLARK: Yes, okay, let's see, so, now, we're onto what, 11-3, which is level of -- oh, no, we talked about that.

Outliers and thresholds, I guess. So, guidelines, not specifications. I think they did the same thing, here.

Let's see, provider reports, the observed episode cost Winsorized at the second and 98th percentile.

Claim line outliers are not removed, and the use of risk adjusted results are intended to correct for extreme outliers.

So, I guess a question here is then, when you're talking about Winsorizing these outliers, is that at the -- that is at

the episode level? Is that -- it's not at the claim level, right?

DR. WEISS: I'm sorry, Winsorizing happens at two levels.

For hospitalization, it happens at the 99th percentile, for episode, and then it happens again at the episode level, at the second to 98th percentile.

MS. CLARK: Okay, sample size requirements, no specifications there. Anyone have comments on that? Same ones? Same comments?

What did we say about sample size, before? I don't remember.

CO-CHAIR CURTIS: I think it was a little different kettle of fish, because it was at the hospital -- it was at the physician level, previously.

MS. CLARK: Okay.

CO-CHAIR CURTIS: And that's, I think, a different -- you know, I mean, the same sample size issues are -- so, it would be

nice to have a --

MS. CLARK: A number?

CO-CHAIR CURTIS: -- some sort of assessment, a threshold, with some empiric evidence to back up why they selected that, that threshold.

However, there isn't that threshold met for the outcomes measures, where they just arbitrarily chose 25.

MS. CLARK: Okay, all right, and --

DR. WEINTRAUB: It certainly can be done, per the comment from Carlos, yesterday, some modeling -- not modeling, simulation exercise could help in the sample size.

MS. CLARK: Let's see, then the last one in this section, defining, bench marking or comparative estimates.

So, these, again, are provider level summaries, and they go through the method of calculating the cost at the provider

level, looking at observed to expected cost ratio.

Now, I think there might be a cut and paste problem here, too, though, because it's kind of mixing physician attribution in with this hospital.

So, this is --

DR. WEINTRAUB: You're at the page --

MS. CLARK: -- needs to be cleaned up, here.

DR. WEINTRAUB: The top of page 30 is what you're talking about?

MS. CLARK: Yes.

DR. WEINTRAUB: Yes, I see it.

MS. CLARK: So, that would be the comment here, I guess.

DR. WEISS: I'm sorry, are you talking about 12.2?

MS. CLARK: This is -- no, this is 11.6, and it's the top of page 30. It's the last paragraph in this, in 11.6.

So, it's talking about -- the very last -- let's see, where is that?

CO-CHAIR ROSENZWEIG: Provider summary reports.

MS. CLARK: Yes, I mean, you're talking about, for example, if the provider for which the summary statistics are being calculated, as a general internist, and it's the --

DR. WEISS: Okay, yes, thank you. We can fix that.

MS. CLARK: Okay, yes.

CO-CHAIR CURTIS: And similar for the sample report that you provided, it's at the physician level.

DR. WEISS: Yes, we can make that change, too, sorry.

MS. CLARK: Okay, so, should we go onto the reliability piece, or go ahead and vote on this part?

CO-CHAIR CURTIS: I think we should, again, go with 2A1 and 2B1 --

MS. CLARK: Two-B1?

CO-CHAIR CURTIS: -- voting, which considers these criteria S11 to S11.6.

MS. CLARK: Okay.

CO-CHAIR CURTIS: So, I don't know if you can put up the table of previous votes, it would be kind of useful.

I think we've identified unique aspects of this one, that are worth consideration, or different than the previous measures. But there is a lot of overlap, as well.

So, for 2A1, the measure is well defined and precisely specified, so that it can be implemented consistently within and across organizations.

Before we put up the vote, any other further comments or general summary comments? Mary Ann, do you want to tell us what your thoughts are on this?

MS. CLARK: Well, I mean, I think, you know, we've discussed, there are quite a

few issues that need to be corrected here.

So, I mean, I don't know that we can go forward, you know, I would either say medium or low, on this measure.

CO-CHAIR CURTIS: Okay, let's go ahead and vote on that.

So, three medium and five low.

And then for 2B1, the measure specifications are consistent with the evidence presented to support the focus of measurement under criteria in 1B. The measure is specified to capture the most inclusive target population indicated by the evidence, and exclusions are supported by the evidence.

I'll editorialize a little bit, that I think the exclusion of the SNFs and in-hospital mortalities makes me more concerned about this than I was on the previous measure. Are there any other comments?

MS. CLARK: No, I think that's -- I agree.

DR. WEINTRAUB: Yes, I agree, as

well. I think those are problematic.

CO-CHAIR CURTIS: Okay, go ahead and vote.

And so, then you want to go to reliability and validity? I'm sorry, it was eight low, is that right?

MS. CLARK: See if we can find that. So, that's on page 31.

Okay, so, here is where there are -- they're doing the testing on the Thomson Reuter's database, as well, they say, a sample of CMS Medicare data.

So, I believe the same type of testing was used on the Thomson Reuter's database, and then they talk about testing on the Medicare database, a sample of 100 percent of the Medicare population in 12 metropolitan areas, and I guess that's kind of the one I was most interested in.

They said it was necessary to make some modifications to the analytic methodologies, in the Medicare analysis,

Medicare testing, and I don't know that those are really specified, though.

CO-CHAIR CURTIS: Isn't that probably around the pharmacy, absence of the Part D, or is that separate from that?

MS. CLARK: They did that, but they also did something with costing, too, which I think they changed a costing method.

DR. WEINTRAUB: They say what the modifications are on the top of page 33.

MS. CLARK: Yes, developing a new set of prices, to be applied to individual services and hospitalizations.

So, that's a mystery. You know, what does that mean?

DR. WEISS: I can clarify that. Sorry for the lack of clarity in here.

There wasn't a one-to-one cross-mark from our average -- our standardized price data that we created from the Thomson data sets, to what we had in the Medicare data sets.

So, we just created a new standardized price table.

MS. CLARK: How?

DR. WEISS: Same way that we did it for the Thomson data, which we talked about a little bit yesterday, where at the hospital -- we do it at the hospitalization level, and then we do it for other events at the CPT or procedure code and modifier level, creating an average cost for those events, and then applying that every time we see that event in the data set.

MS. CLARK: In this case, I'm wondering why, you know, the Medicare methodologies weren't used.

You know, there is some specific methods that CMS uses to cost out services, which could have been employed here. Just an idea.

So, and also then the SNF claims were dropped, as well, or was that -- those were not included here, either, right?

DR. WEISS: Right, same exclusion criteria.

MS. CLARK: And I see that also there were -- analyses were dropped of resource use -- well, by individual provider, that wasn't part of this, but -- and provider specialty, but you're talking also, about dropping analysis of individual hospitals, as well, but hospitals are in the Medicare data. So, why would those have been dropped?

DR. WEISS: Well, this statement, we were, at the time, investigating whether or not this was -- this measure was going to be able to be attributable at the team level, within a facility, so that we could identify the group of providers that were providing care to the patients.

We attempted to do that. We found that we were unsuccessful in doing that, both in the Thomson data and in the Medicare data.

I don't know how many hospital identifiers were missing, when we went to

hospital level attrition in the Medicare data. That is something that we can provide additional information on.

MS. CLARK: Well, you should have had all the hospital identifiers, because that is how they get paid. So, okay.

So, testing results, market scan testing, do we want to go to the slides that present the results?

Let's see, where are those located? Those are a separate slide?

MS. TURBYVILLE: They're in the PDF.

MS. CLARK: Yes.

MS. TURBYVILLE: They are the fourth bookmark.

MS. CLARK: Fourth bookmark, scientific acceptability attachment? Okay, same slides?

DR. WEINTRAUB: Same sort of orientation slides we've seen before.

MS. CLARK: Anything new here?

DR. WEINTRAUB: Actually, there is.

MS. CLARK: Okay.

DR. WEINTRAUB: The distribution of costs is not as problematic as we've seen before.

MS. CLARK: In terms of the -- the distribution, in terms of what? The related, AMI related services, or the non-related services, or just in general?

DR. WEINTRAUB: Actually, I don't think they lay them out quite the same way, unless I'm missing something.

MS. CLARK: Yes.

DR. WEINTRAUB: So, they're not related to four. They had, within that, they had related and non-related. I don't see that distinction.

MS. CLARK: They're separate slides, it looks like.

DR. WEINTRAUB: And one thing that's going to make that analysis a little

bit easier, if you go to slide 15, which is 63 in the PDF, you see the overall distribution and cost is not as skewed.

MS. CLARK: Well, yes, because everybody had a hospitalization --

DR. WEINTRAUB: Yes, right. So, it does make it a little bit easier.

But the outpatient costs, on the other hand, are -- it's a smaller piece, and there is more skew.

MS. CLARK: Yes, well, that's, again, a question, because I mean, out-patient facility cost, where is that even coming from? I don't know. Is that --

DR. WEINTRAUB: I agree.

MS. CLARK: That wasn't one of the resource use specifications. Is it a BETOS category? I don't know.

CO-CHAIR CURTIS: Could the developer just clarify that?

DR. WEISS: Yes, we categorized it, based on the category -- we know, for our

-- I'll say this with a caveat that we know, in the data set, we have a difficulty identifying all of the outpatient facility costs that occur, and appropriately grouping them.

We don't have a problem finding the costs that occur, but we have a problem appropriately putting them in this cost bucket.

So, while some of the facilities, you may be able to do a good job of identifying the facility costs, others may show up in the procedures bucket, or potentially in the physician services bucket, under E&M.

So, while we are relatively confident we're capturing the full spectrum of costs, there may be some mis-classification, in terms of these descriptors.

DR. WEINTRAUB: The one I'm a little concerned about is other services, quite skewed. It's only two percent of the

overall, but it's zero from the 75th percentile, and then it goes up to $2,500. What is that?

DR. WEISS: It's a mix of stuff. I mean, it's a bucket that captures lots of different things.

I mean, I can get you the BETOS list, to show you the groups there. The problem is, it may be capturing some of our outpatient facility costs, and again, I'm going to fully admit that we have some problem in appropriately categorizing costs into all of these buckets.

The biggest thing here is that 81 percent of all costs are on the inpatient side.

CO-CHAIR CURTIS: So, I think, just -- it's a similar approach to reliability and validation that we've seen, you know, the specific data, probably a little bit less problematic, given that we have the inpatient admission for all patients.

MS. CLARK: Right.

CO-CHAIR CURTIS: It's less skewed by subsequent admissions.

But you know, partly in the interest of time, I kind of want to make sure that we're focusing on the differences, as opposed to going through entirely de novo.

MS. CLARK: Okay, so, let's see, what page are we on, there?

So, that was reliability testing, and validity, right?

So, are we ready to vote on those, or do we need more review?

MS. TURBYVILLE: You could also look at any of the other commenters on this sheet, that you have, maybe, and --

MS. CLARK: Sure.

MS. TURBYVILLE: For 1570, it would have -- I think -- though, you may have been the only one þ-

MS. CLARK: Was I the only one --

MS. TURBYVILLE: Yes.

MS. CLARK: Yes, I don't think I have anything to add to myself. I mean --

MS. TURBYVILLE: Well, you could have changed your mind.

MS. CLARK: Yes.

MS. TURBYVILLE: Yes, let's see.

CO-CHAIR CURTIS: I think if everyone is comfortable with what we're doing, I think we should go ahead and vote.

We have the -- on the screen, the comparison of the two tables, of the prior measures, and how we voted on reliability and validity in those cases.

We would start with 2A2, reliability testing demonstrates that the results are repeatable, producing the same result a high proportion of the time when assessed in the same population and that the measure score is precise.

And in looking at that, we have definitely directed from 2A2 in the first one, to the second, where we would move from mostly

moderates to all lows.

DR. HWONG: I think it had much more to do with the actual measure construct, right?

CO-CHAIR CURTIS: I agree, agreed.

DR. HWONG: Yes, right.

CO-CHAIR CURTIS: So, I think that --

DR. HWONG: Reflective of that.

CO-CHAIR CURTIS: Yes, and then specifically --

DR. HWONG: As opposed to the approach.

CO-CHAIR CURTIS: -- that missing of the -- or the combination of the -- in relation to the specification.

So, let's go ahead and vote on that one, and we're -- okay, there it is, okay.

MS. WILBON: Two moderate -- I'm sorry, seven moderate and -- seven moderate and one low.

CO-CHAIR CURTIS: So, 2B2, validity testing demonstrates the measure data elements are correct and/or the measure score correctly reflects cost of care and resources provided, adequately distinguishing higher and lower cost or resource use. And it was again moderate for 2B2 in the first, and eight on the low.

I would say that the -- in my opinion, this is more likely, or closer to the first measure that we would be within the second.

That being said, there are also new caveats of the issues that we raised. So, go ahead and vote on that.

Four, sorry, six moderate and two low. So, for 2B3, exclusions are supported by the clinical evidence, otherwise, they are supported by evidence of sufficient frequency, but mainly, focusing on the measure specifications and how to specify the exclusions.

So, in the previous ones, for the first measure, it was a range around moderate and the second, it was five lows and two mediums.

MS. TURBYVILLE: Is this one influenced by the discharge question?

CO-CHAIR CURTIS: In my opinion, it's influenced by the exclusion of in-hospital mortality and the discharge to SNF.

That's eight low.

For 2b4, outcome measure, more resource use when indicated, evidence-based risk adjustment strategy is specified, based on clinical factors, and for that we were fairly consistent with the range around moderate, for both measures, and I think it's consistent with that.

Go ahead and vote. So, that's eight moderate.

For 2b5, data analysis demonstrates that methods for scoring at the -- the analysis has specified measure allowed

for identification of statistically significant and practically clinically meaningful differences in performance.

For the previous ones, we had mainly -- well, insufficient on the second measure, post-revascularization, and more moderate to low on the first measure, the 31 to 365.

Mary Ann, do you have a thought as to kind of where this would fall?

MS. CLARK: You know, the testing they did was -- or creating a score, this one -- this one was at the hospital level.

So, you know, I'm not -- I wish we had our statistician here, but so, you know, I'd say this was probably more of a moderate to me.

DR. WEINTRAUB: So, the testing shows some range to it, but they really don't have model characteristics. They haven't done validation. They haven't done calibration, I mean, they're not -- they haven't tested in

external data sources.

I mean, it seems to me a lot of them could be done.

CO-CHAIR CURTIS: Okay, let's go ahead and vote on that. That is three moderate and five low.

Then, 2b6 is actually I think relevant for this, because of the -- potentially, because of the CMS data, but let me think about that. Probably not, because it's the same administrative data.

DR. HWONG: Yes, I don't think they're saying that there is options for use.

CO-CHAIR CURTIS: So, I don't think we'll vote on that, and for the disparities, similar to the other ones, we do not vote. So, keep that consistent.

DR. WEINTRAUB: We're not going to vote?

CO-CHAIR CURTIS: We're not going to vote on those two, for reliability, not applicable for this measure.

So, I think what we discussed yesterday was, in the interest of time, for usability and both feasibility we would anticipate that this would have a similar score, basis of similar issues.

Eventually, this will be voted on formally by the TAP, or those -- that information will be captured, as we anticipated for all these measures, there will be the opportunity for a re-vote, via a SurveyMonkey or other device.

So that the formal opinion of the TAP will be captured, but we will save a few minutes, at least, in terms of going through and getting the vote and the delays with the reply key.

So, just to clarify that for the measure developer, this will be officially done. We would anticipate that it would be the same voting at this time.

There were some efficiencies captured in this. I think we should continue

our plan, move to the ABMS diabetic measure, and then go to the Ingenix measures to follow that.

But let me ask you, should we take a break now, or do you want to --

CO-CHAIR ROSENZWEIG: Yes, let's take a --

CO-CHAIR CURTIS: Maybe a five minute break?

CO-CHAIR ROSENZWEIG: Five minute break just for restrooms.

(Whereupon, the above-entitled matter went off the record at 10:09 a.m. and resumed at 10:15 a.m.)

CO-CHAIR CURTIS: Why don't we go ahead and reconvene?

CO-CHAIR ROSENZWEIG: Yes, in the interest of time, let's get started.

We're now considering review number 1576, I believe, episodes of care for patients with diabetes over a one-year period, and Brenda Parker will be the -- is the

primary reviewer.

Do we have any comments by the measure developer first, before we start an introduction?

DR. WEISS: I'll just give you a very brief introduction.

The measure is developed in the same manner as the other measures that you reviewed from us.

The intent of this measure was to focus on patients that were not newly diagnosed with diabetes nor were at the end stage part of their disease.

We were trying to focus on a group of patients that our work group sort of termed in the management phase of diabetes, and we did this by identifying homogenous patients, in an attempt to capture all of their diabetes=related resource use over a one-year period.

CO-CHAIR ROSENZWEIG: Okay, thank you very much. Brenda, would you like to

start?

MS. PARKER: Sure, thank you. Thank you to the measure developer for the brief overview.

So, with that, we will jump right into the importance to measure and report.

Regarding impact, there is, in my opinion, sufficient evidence that the measure developer has provided, in terms of high impact, regarding the epidemiology of diabetes, as well as some of the care considerations and the economic consequences of diabetes, in terms of co-morbidities.

So, for that, I think that they did a great job, there. Also, a note to them for identifying the IOM as ranking this as a top 20 priority, in general.

In terms of opportunity for improvement, regarding variation across providers, as well as disparities in population groups, it was definitely sufficient evidence to support that practice

variation does exist, as well as racial disparities, within diabetes care.

However, racial disparities were really the only variation that were targeted. There were no other discussions of other population groups where disparities may exist, as far as age, gender, socio-economic status. So, I think that was a deficiency there, within that section, because again, it did concentrate primarily on race, which, there is a lot of data to support that focus, but I think, you know, in terms of being a well rounded measure, you should probably attempt to address multiple disparities.

With regards to -- sorry, I got ahead of myself here.

With regards to the purpose and the intent, and that is on -- starts on page six, my apologies for not keeping everyone along as we go, with the bulk of that being at the top of page seven.

The intent of the measure was

clear, however, after going through the AMI intent, or measure this morning, I'm wondering if -- because the question of re-admission came to mind, and why this was included, as it wasn't included in the previous literature, as it being a large issue in patients with diabetes.

So, I'm wondering if this was, because essentially, it's the exact same language from the AMI, so, I'm wondering if that may have been a copy and paste, or if the measure developer could provide some evidence or support of why, in addition to resource use, re-admissions were mentioned specifically within this section.

Would the measure developer care to comment on that?

DR. WEISS: I'm sorry, can you point me to the page? I'm trying to find it.

MS. PARKER: Sure, it's on page seven, at the top of the PDF, and there is, I believe it's in the second sentence, provide

efficient care, third line, by examining both the resource use, as well as the re-admission rates.

DR. WEISS: Yes, this partially would be a problem with the copy and paste, and I apologize, I thought when we were looking at this, hospitalization -- and re-admission is probably the wrong term; it's more hospitalization ends up being an important cost driver within this episode. So, we can clarify this language.

MS. PARKER: Well, and so, is there really the need to include something more specific when later on, in the construct, you mention that in-patient hospitalization is included in there?

So, I would think that that would fall under the blanket umbrella of resource use.

DR. WEISS: Fair enough.

CO-CHAIR CURTIS: And I think there is a fair intent here, to compare the

relative research used by different providers to examine patterns in diabetes.

MS. PARKER: Right.

CO-CHAIR CURTIS: And compare the healthcare costs, so, I think --

MS. PARKER: Yes, and in my notes --

CO-CHAIR CURTIS: So, I don't think we should cut down too much --

MS. PARKER: -- it's fine, yes, it's fine, just wondering if I missed something, or if there is something that the measure developer cared to elaborate on.

And then finally, within the importance to measure and report category, evaluation of the resource-use categories, which seemed to be relatively consistent with some of the other measures that have been -- and I'm trying to get to that section, my apologies.

Trying to get down there. It seemed to be consistent with some of the other

measures, given that this was in -- a chronic condition.

One thing that came to mind, perhaps that relates to this, sort of is within the care setting, and I didn't find those individual care settings. That seemed to be a very short list.

So, I'm not sure if that's something that, as mentioned earlier, the care setting categories that are provided, the taxonomy there, if those are more of the broad categories, and a lot of the sub-categories may roll into that, because there were only three care settings identified: either hospital, primary care or pharmacy, I think.

And there seemed to be a lot more granularity in some of the other measures, or at least more granularity.

So, that may be an NQF question, or a measure-developer question, but I think that gets at the resource-use categories.

My question is not clear, is it,

because you're looking at me.

MS. TURBYVILLE: I just want to make sure, so, is it that you felt that perhaps, they weren't inclusive enough of the resource-use categories for this measure?

MS. PARKER: The resource-use categories themselves, as indicated in S96 --

MS. TURBYVILLE: Okay.

MS. PARKER: -- seem to be adequate, but they don't match up to what I expected to see in the care setting.

So, just trying to reconcile that, while we're at this point.

MS. TURBYVILLE: Okay.

CO-CHAIR CURTIS: So, is that a function of, again, they're sort of doing diabetes-related resource use?

MS. PARKER: It doesn't seem to be that it would be a function of that. It's merely a function of what they chose, or maybe what they had an option to choose from.

CO-CHAIR CURTIS: And are you

concerned that there are specific areas that --

MS. PARKER: That may be missing in the care setting.

CO-CHAIR CURTIS: So, which ones do you -- well, maybe we'll come back to that.

MS. PARKER: Sure, but that doesn't mean you can look at --

MS. TURBYVILLE: Well, we'll have to look at our list and --

MS. PARKER: Yes, perfect and we'll get to that later.

MS. TURBYVILLE: Okay.

MS. PARKER: So, that concludes kind of my overall cursory review of the importance to measure and report.

So, if there are no further questions or no further comments from the measure developer, I think we could go ahead and vote.

CO-CHAIR ROSENZWEIG: Sure, any comments or questions? All right, so, let's

vote on 1a.

MS. PARKER: And within this category, I'm happy to kind of give my input in how I ranked this, as I was reviewing it.

Again, 1a is the importance of it and the impact and again, that provides sufficient evidence. So, I ranked that actually as high.

CO-CHAIR ROSENZWEIG: All right, eight, high, and it's in the right color, it fits in this time.

MS. PARKER: That's helpful.

CO-CHAIR ROSENZWEIG: All right, and then the second, 1b is demonstration of resource use or cost problems and opportunity for improvement.

MS. PARKER: Sure, and with that one, as I mentioned previously, I think that they did a good job, as far as practice variation.

But, and Jamie, you could probably comment to this, if there is sufficient data,

as far as disparities regarding age or gender or socio-economic status, because I think there was a real deficiency there in the development of their case.

So, I ranked that as moderate, because again, they did a great job on practice variation and the literature of the racial disparities was thorough. I just think they kind of missed an opportunity.

CO-CHAIR ROSENZWEIG: And there were also opportunities for resource use that they didn't mention --

MS. PARKER: Sure.

CO-CHAIR ROSENZWEIG: -- related to actual appropriate resource use --

MS. PARKER: Right, yes.

CO-CHAIR ROSENZWEIG: So, there were a lot of -- for instance, absence of eye exams and the various other things.

So, there is -- we're not just talking about saving money, here, we're also talking about appropriately using resources.

MS. PARKER: Exactly.

CO-CHAIR ROSENZWEIG: And there are clear disparities that have been identified.

MS. PARKER: I agree. So, again, I chose to kind of rank this as moderate.

CO-CHAIR ROSENZWEIG: Two high and six moderate, okay. Then the third is 1c.

The purpose objective of the resource-use measure and the construct for resource use are clearly described.

MS. PARKER: And for what it's worth, because of the re-admission piece, it confused me a little when I was reviewing it, and I ranked this moderate.

I'll leave it up to the panel, to decide, you know, based on the conversation here, what they would like to do with that.

CO-CHAIR ROSENZWEIG: Two high, five moderate and one low.

Okay, and then the fourth one, 1b, the resource-use category that are included in

the resource measure consistent with and represented of the conceptual construct, represented by the measure.

So, they want to make sure that the categories that are being used are coherent and consistent with the purpose of the measure.

MS. PARKER: Sure, and this specific resource-use categories that they did identify are comprehensive and specific, regardless of the care setting.

CO-CHAIR ROSENZWEIG: Okay.

MS. PARKER: I ranked that high.

MS. TURBYVILLE: Can I ask one clarifying question for our notes?

For the purpose and intent, we had a little bit, you know, the moderates and the lows.

I just want to make sure that we captured the feedback, which is that they're pointing to some resource use and not other, which others, which may signal that some are

more important and others ought to -- other than just being broad, and thinking about resource use and measurement of diabetes.

For example, re-admissions was mentioned, and it didn't really seem to --

MS. PARKER: I think the two statements you have are crossing. I think the initial statement you had relates more to the 1b, so, regarding variation and practices.

MS. TURBYVILLE: Okay.

MS. PARKER: The re-admission piece was perhaps, a copy and paste error or an element that the measure developer said that they would modify and/or remove all together.

MS. TURBYVILLE: Sounds good, okay.

CO-CHAIR ROSENZWEIG: Okay, let's move on to scientific acceptability.

MS. PARKER: So, if it's okay with the panel, for scientific acceptability of the measure properties, what I did in reviewing

this again last night was, I kind of called out the things that were similar to the other proposed measures, and I'd like to just review those.

So, it may seem like it's jumping around a bit, but I think it's helpful to go ahead and review what we've kind of already processed, if that is acceptable.

CO-CHAIR ROSENZWEIG: Sounds good.

MS. PARKER: Great. So, the general approach is the same, as far as establishing a working group to kind of weigh in on these different measures.

The data protocol itself, again, is very consistent, as far as they do not recommend imputation of missing data, that only closed claims, or those that have been paid are utilized and that the quantity values for resource use and the frequency are set to one, when missing an order, to capture costs.

The data type is administrative, as are all of the proposed measures. The co-

morbidity risk adjustment used is the HCC, which they've proposed in previous measures.

Costing, again, uses the NCQA standardized price tables and their modifications to such tables.

And so, I think all of those are pretty consistent, and what they've presented is consistent, but also, the feedback in the context of diabetes is consistent, as well, in my opinion.

CO-CHAIR ROSENZWEIG: Okay.

MS. PARKER: As far as the relation to -- let me get the overall, broad category here.

The detail attribution peer group, outliers, table size, bench marking, that grouping, the attribution is 70/30.

This is a chronic disease state, as far as, you know, most of the care, in my experience, and anyone can weigh in, is that, you know, a primary care provider does typically manage patients with diabetes and

refers those on who need further follow up or have more severe disease.

So, it makes sense that the majority of this would be attributed to where most of the care takes place, versus with some of the other events in cardiology, where it may be that it's referred -- the patient is referred to a cardiologist, and this may increase costs, and therefore, the attribution may be a little skewed, in terms of that.

I think the attribution method here, in the context of diabetes, makes sense.

Anyone have any comments?

CO-CHAIR ROSENZWEIG: Yes, I would just propose some caveats, and that is that patients with more complex and more expensive diabetes tend to be referred to endocrinologists for care, if they need to -- usually, those people who are on insulin or are on multiple daily injections and who have multiple complications.

So, the attribution of those

patients is more likely to be to the endocrinologist, since once they refer to an endocrinologist, it's usually not for a single visit, it's for multiple visits.

And whereas, the attribution for the other patients who might be -- might have lower costs, might stay with the primary care doc.

One of the reasons they're actually referred to an endocrinologist is when they need more care.

MS. PARKER: Absolutely.

CO-CHAIR ROSENZWEIG: So, there may be issues related to case-mix adjustment that this particular model may not fully take into account.

MS. PARKER: And if the measure developer could address that, I think that would be very appropriate -- an appropriate request.

They do mention and acknowledge that the more severe patients will be referred

to an endocrinologist. So, they do note that that does occur.

CO-CHAIR CURTIS: That seems like it should have been covered at least at -- if the primary comparisons are within peer group, and the peer groups are appropriately defined, then it would be less of an issue.

But it has to do with, I guess, how the measure is used at the end of the day.

CO-CHAIR ROSENZWEIG: Correct, it's just that, yes, diabetologists are not -- it's interesting, but there are some diabetologists who are not endocrinologists.

And so, there is a -- it's a bit of a -- it's not as clear as the attribution issues that apply to chronic CAD.

MS. PARKER: Sure. The level of analysis is at the individual clinician level, or proposed that way.

Winsorization, as in the other proposals takes place, and there are no sample size recommendations and finally, the bench

marking is, in terms of the provider summaries.

I had no objection to it being a provider level, but I wanted to throw that to the actual providers to weigh in as that's not really my area of expertise.

DR. MARWICK: How is the attribution made?

If somebody, for example, has a coincidental identification of diabetes in the midst of another problem, is the diabetes attributed -- presumably, the diabetes is attributed to the person looking after the other problem? I just see that as a potential issue, here.

MS. PARKER: And if it helps, we could go ahead and go through the clinical framework, so, that you know, kind of, how the population is identified, or if the measure developer would like to kind of comment, on that request, at this time.

DR. WEISS: Sure, I mean, the

attribution logic function around the E&M codes that have a diagnostic code but groups different episodes.

So, we identify all of the physicians and -- the provider interactions with an E&M code that have eligible ICD9 codes for this episode, and then, make the attribution rules, based on the proportion of those codes that are -- the proportion of those digits that are acting within a provider or multiple providers.

So, in the example, if the person has another problem, and they're going to a cardiologist, for example, and the cardiologist also includes a diabetes code on that claim, there is a possibility that the cardiologist would be the one that is attributed in the episode.

MS. PARKER: Does that clarify the --

DR. MARWICK: Yes, it does. It was kind of what I was afraid of.

So, you know, the risk there is that the cardiologist may not be the primary person looking after the diabetes, and it may not be wise to attribute subsequent interactions to them.

MS. PARKER: Sure, and is there a way that this -- this may have been tested, and I just didn't see it, but is there a way, from the measure developer, to give us an idea of how often that happens and/or maybe a proposed approach to how this can kind of be addressed, or minimize it, at best?

DR. WEISS: So, in our testing, cardiology happens to be the sixth most common specialty for which episodes are attributed to, but it's a -- only 3,000 episodes versus family practice, which is 41,000 and internal medicine, which is 33,000.

So, the absolute number being attributed to a cardiologist is much, much lower than other types of providers.

CO-CHAIR ROSENZWEIG: Correct, but

the costs may be much higher, knowing that cardiologists tend to cost more.

DR. WEISS: Understand, and that's why we would not propose comparing episodes attributed to a cardiologist, compared to some -- and episode attributed to a family practice physician.

We'd only want to compare episodes that are attributed to cardiologists with other cardiologist-attributed episodes, those providers.

MS. PARKER: And that seems to be a sufficient and adequate comparison.

DR. MARWICK: I think there are going to be statistical issues there, that's part of the concern, and I wonder if the more sophisticated approach would be to look at, for example, if a patient has had multiple visits, maybe there is a threshold number of visits with the attribution -- with diabetes linked that would be a better means of doing this.

As I currently understand, it will be possible for somebody to see a cardiologist once, and have that listed, and then the cardiologist be linked to that patient in subsequent events.

MS. PARKER: Is that something the measure developer would be willing to do and/or potentially address here?

DR. WEISS: Would we be willing to change our attribution logic? Is that the question?

I think we're well beyond our ability to change our attribution logic right now.

MS. PARKER: Well, I think more so, it's just maybe a proposed -- you know, really, the burden is on you to kind of identify what that appropriate number would be, I believe, or if that makes sense -- and either it does or it doesn't for your proposed measure -- but that's something that only you can kind of decide, if that's something that

makes sense, and if that's something that you could do.

CO-CHAIR CURTIS: I guess to me, it just doesn't -- there's going to be noise in this --

MS. PARKER: Sure.

CO-CHAIR CURTIS: There's going to be mis-classification, but as we said, the sample sizes are going to be extremely low, and I guess if I got that report back, of characterizing the care of my diabetic patients, and I was signing it, I guess it depends on the consequences of it.

But I would ignore it, whereas, I think the people -- the primary care doctors and the endocrinologists, diabetologists, would be the ones who would really focus on it.

So, I'm just -- I wish that it were completely precise, but I think it might be an impossible threshold to set.

MS. PARKER: Well, and I believe

NQF has committed to providing a statement, a caveat statement, regarding the statistical significance or the power behind this.

So, I would just charge the Steering Committee with making sure that that statement is accurate, and make sure that it definitely reflects the intent, because I think statistical significance may not be necessarily the most appropriate terminology.

But I trust the Steering Committee will make the appropriate decision.

DR. WEINTRAUB: I want to go back to your original question about whether we can -- will have adequate power to look at the individual provider level, and in this measure compared to others. I think for most providers, we probably can.

It looks like, first of all, there are lots of patients with diabetes.

MS. PARKER: Sure.

DR. WEINTRAUB: And family docs and internists will see a lot of them.

So, if we believe that the attribution, even if imperfect, is okay, or at least acceptable, probably most of the time, we're going to be okay on power.

The other thing, we're looking at the distribution of costs, it's not as skewed as it is, because we're dealing much more with out-patient care rather than hospitalizations.

MS. PARKER: Sure.

DR. WEINTRAUB: So, most of the time, it's going to work out okay.

I do worry that low-volume providers, or providers that have a couple of hospitalizations, may find all of the sudden, that they're lying outside the boundaries, here and there.

MS. PARKER: And that may be something that the measure developer can consider for the next step in the process, as NQF mentioned, the three-year revisiting.

Maybe by that time, you know, if the measure is approved, or if they have

adequate time to develop an adequate sample size, that they can test and make sure, you know, with this many -- not necessarily number of providers, but as NCQA uses, but maybe number of visits or -- et cetera -- getting back to the general comment of cardiology, but also, kind of the concept of adequate power.

CO-CHAIR ROSENZWEIG: Yes, I think adequate power may be an issue, even though diabetes is a common disease, still represents, you know, in many plans, only about four to five percent of the population, okay, because they exclude the elderly, and that's not enough, necessarily, in a typical primary care practice, or in -- you know, to really necessarily achieve good power.

And this has already come up in quality improvement measures, okay. So, there is no reason why it shouldn't come up here, and physicians with part-time practices as well: that will be even a bigger problem.

And since measures like this might

be used for things like physician tiering, which we already have in Massachusetts, I don't know if you have it in your states, this is an important issue.

MS. PARKER: Absolutely, and that kind of rounds out the -- what was similar. So, I think this was very good context for diabetes.

I also think we land kind of in the same general area, with a few exceptions.

So, now, I'm going to go specifically to the clinical framework, which is specific to diabetes. So, it obviously, is different from the other proposed measures that are specific to cardiology.

And just walking through, I'm going to walk through the clinical framework, before I kind of digest it for you, and for the measure developer, please feel free to jump in and correct my interpretation of what you have here, if that applies.

So, essentially, the way that the

patients are going to be identified, and there are two methods, if you will. The first one is really using the oral medications, in terms of identifying patients. So, there are no age restrictions within the first criterion.

Essentially, patients are required to have at least one out-patient visit with a diagnosis of diabetes, within the first six months of the identification year.

So, again, 24 months is within the document. Within the first year, where the patients are identified, they need to have a diagnosis of diabetes within the first six months, at least one prescription for an oral hypoglycemic medication in the first six months, as well, and at least one diabetes-related resource-use event in the measurement year.

That could be anything from a fill at the pharmacy, a visit to the doctor, even a hospitalization. So, just one measure.

Again, I think similar to one of

the measures yesterday, you know, just making sure that these patients are -- have not left the Earth, I believe is the exact phrase from yesterday. So, that's the first criteria.

The second criteria is looking more -- using insulin as kind of the differentiator.

So, again, within the first six months of the identification year, needs a diagnosis of diabetes. No oral hypoglycemic medication in the first six months of the year, rather, they would have one insulin claim in the first six months of the year, and there is an age restriction here, and it's restricted to those patients who are 30 years and older during the identification year, and at least one diabetes resource-use event in the year of measurement. That's the inclusion criteria.

So, I'll go ahead and go through the exclusion criteria, then maybe we can discuss both, or does it make more sense to go

ahead and discuss the inclusion criteria and concerns that I have?

DR. HWONG: I was wondering if we could --

CO-CHAIR ROSENZWEIG: Let's just -- go ahead.

DR. HWONG: -- pause for a second and -- go ahead.

CO-CHAIR ROSENZWEIG: No, go ahead.

DR. HWONG: I was thinking, you know, I have a couple sort of questions --

MS. PARKER: Okay.

DR. HWONG: -- in terms of the inclusion, and sort of, how they've -- because I'd expect it this way, right?

So, one of the -- yes, I guess I would love to get some clarification on, you know, the alternate path to get into this measure, where -- yes, if you're an insulin user, so, I'm assuming that's probably going to help you identify a lot of your Type 1

diabetics.

But why is this 30 -- why for that path, do you have to be 30 or older?

MS. PARKER: Which, I would think is -- I was thinking the same thing, until I got to the age 30, and that's counter-intuitive, because Type 1s, you know, are -- could be younger, usually more healthy, not your typical Type 2s.

So, I thought that was a -- and there is no real rationale, in my mind, of why 30 is an appropriate cut off.

So, I was hoping that, again, either the practicing docs here could help or the measure developer could weigh in.

CO-CHAIR ROSENZWEIG: Yes, I was concerned about this, too. Could the measure developer comment on this?

DR. WEISS: Yes, sure. Our clinical work group pushed this forward, the identification, the focus of this measure being on patients with Type 2 diabetes.

We realized that through our specification criteria, for inclusion, there will be some Type 1 diabetics that enter into our population.

The second inclusion criteria was an effort to identify the insulin-only Type 2 diabetics, and that's why the age restriction was placed on the second pathway.

MS. PARKER: And so, here are my thoughts. I automatically identified that there was no separation of Type 1 and Type 2. You just talk about diabetes in terms of the measurements.

So, that is not clear, and could be made more clear, and there are specific ICD9 codes that help with that diagnosis, assuming that they are not mis-classified.

So, I think there could be a combination, but you know, helping to clarify that earlier on, probably would have taken care of this concern, at least as it stands now, of why that was the case.

But I'm still not sure that if -- that you would want -- I don't -- I wrote down two things, as far as this, that relate, that Type 1 and Type 2 could potentially be stratification, because they are different to your point, Type 2 being, I think, 90 to 95 percent of the population with diabetes.

And then also, the exclusion of the newly diagnosed, as well as the end stage, again, being kind of stratification measures, because they still all have diabetes, and we still all want to know about their resource use, and match them to quality, so, that we really understand the efficiencies of care in diabetes.

So, I'm not sure that the measure is specific to Type 2, I don't know if that was a request or that is just the general consensus, that that's where we need it. But if we have the opportunity to include some of these other perspectives in care, I think that it is wise to take advantage of that

opportunity without a lot more effort.

CO-CHAIR ROSENZWEIG: Yes, I would just comment, if indeed, the focus then is on Type 2 diabetes, then it should be included in the title of the measure very specifically, if actually, you want to exclude Type 1 patients that are known.

Now, with respect to -- I do understand, though, that there are differences in the various coding for Type 1 and Type 2, but those are often misused, and I think that's the rationale --

MS. PARKER: Absolutely.

CO-CHAIR ROSENZWEIG: -- for their not using them that specifically.

MS. PARKER: Sure.

CO-CHAIR ROSENZWEIG: Because it's extremely common for physicians, once a patient is on insulin, to classify them as Type 1 --

MS. PARKER: Type 1.

CO-CHAIR ROSENZWEIG: -- even

though they may not really be Type 1.

MS. PARKER: Sure.

CO-CHAIR ROSENZWEIG: Nevertheless, I think it's -- I think that if the focus really is suppose to be Type 1, and you're not -- you don't want to have to include Type 2 in this population -- excuse me, Type 2, and you don't want to have to include Type 1 in the population, then it definitely should be part of the actual title of this protocol.

DR. HWONG: Right, and I would say even beyond the actual title, that if this is truly the intent, and that's fine, you know, if the measure developer wants to submit it to be sort of Type 2 diabetes, but just as Brenda has pointed out, there are a lot of other types of criteria you can put in there, to just try and be more stringent.

I understand that there is, you know, some problems with this mis-coding, but I think, you know, there are -- again, sort

of, if that really is the focus, I think you could spend -- a measure developer could spend a little bit more time to try and tighten that up.

CO-CHAIR ROSENZWEIG: Exactly. Now, the other issue is that they don't include all of the various medications.

MS. PARKER: And I was actually getting to that lower down. So, I think that's a great point.

CO-CHAIR ROSENZWEIG: Now, I suppose, for the purpose of -- now, there are two issues.

For the purpose of actually looking at cost, there is -- you need to include all of them, but for the purpose of looking for -- to -- for determining the denominator, I understand why they wouldn't put metformin in, as we --

MS. PARKER: Sure.

CO-CHAIR ROSENZWEIG: -- discussed in previous protocols.

But they have left out a whole number of other medications that are used for treatment of Type 2 diabetes.

MS. PARKER: And I actually had some of my bias from my previous work as, you know, DPP4s, it's a new class, but it's a new class that's being used a lot.

I mean, Januvia is being used quite a bit. So, that's a great point, and I actually have that listed more further down.

MS. CLARK: I just had a question.

CO-CHAIR ROSENZWEIG: In addition, injectable non-insulin medications, like --

MS. PARKER: Yes, Byetta.

CO-CHAIR ROSENZWEIG: -- like sitagliptin, yes, Byetta, Onglyza.

MS. PARKER: Yes, exactly.

CO-CHAIR ROSENZWEIG: Victoza, excuse me.

MS. PARKER: Victoza, yes.

CO-CHAIR ROSENZWEIG: Yes.

MS. CLARK: I just had a question

about the criterion of a diagnosis of diabetes within the first six months of the identification year. What does that mean?

MS. PARKER: It was kind of odd, that it was six months, but --

MS. CLARK: Yes.

MS. PARKER: -- maybe that's again, due to the time frame of having enough claims following the identification. I don't know if anyone else can --

DR. HWONG: I think their mention of the identification year is the year previous -- prior to the measurement, if I'm not mistaken.

MS. PARKER: Right, but why the six months, rather than the full year?

MS. CLARK: Yes.

DR. HWONG: That is a good question.

MS. PARKER: I mean, that may be --

MS. CLARK: I would have

identified patients in the measurement year with diabetes, and then looked back to see -- if you're trying to identify a person that is constantly managed, then you look back in the previous year to identify somebody that, if they had another claim, back then.

MS. PARKER: Well, then it may be to rule out, kind of that new -- that new diagnosis.

MS. CLARK: Well, that is what I'm saying --

CO-CHAIR ROSENZWEIG: They're trying to exclude newly diagnosed patients with diabetes, because they're -- there are a lot of additional costs that occur with newly diagnosed patients, that don't occur in -- subsequently.

MS. CLARK: It just seems like an odd way to do it.

MS. PARKER: Does it make sense to exclude them all together, or to stratify by them?

I mean, to the point yesterday, stratification is a measure used to separate groups that have sort of the same outcomes, but may be different based on that stratification measure, which would be newly diagnosed versus more of the management-based --

DR. HWONG: Yes, I actually thought that might be a good thing.

Like, when I sort of step back and looked at this measure, and I think it's, you know, again, about diabetes management, and I understand, you're trying -- I understand, measure developer, you know, is trying to sort of create this very homogenous group, if you're going to compare across providers, et cetera, to this sort of like, you know, ongoing management of diabetes.

So, but I -- you know, I sort of think about all the individuals that are sort of newly diagnosed, right, and I understand that, you know, costs could be, you know,

potentially higher for these individuals, but it's sort of like without a companion resource-use measure about this group, I think that in some ways, it's -- you know, I would like to see that, right.

If I'm thinking about sort of overall management or care of diabetics, you know, I think that is sort of an important group.

So, you know, since you're saying -- I understand, they're trying to do this sort of homogenous area, but like, you know, something for consideration, I think it might be -- it would have been interesting if they could include those, you know, newly diagnosed, and then, you know, stratify on them.

But like I said, because I could imagine down the road, you can have a whole bunch of resource measures where you're getting to sort of these narrower and narrower, sort of specific --

MS. PARKER: Sure.

DR. HWONG: -- groups, and then that sort of leaves, you know -- in terms of, I think, what it's actually trying to tell you, I think could be, you know, more limited.

DR. REEDER: Are the diagnostic tests and the time involved in creating a new diagnosis for a patient on diabetes, such that there -- it's a long time span, or that the costs are high enough that this particular group, maybe by the developer, was considered an outlier and rightly so excluded?

CO-CHAIR ROSENZWEIG: I think that --

DR. REEDER: I don't know, I'm asking.

CO-CHAIR ROSENZWEIG: I don't recall reading specifically the rationale, but -- in here. Brenda, I don't know, you may --

MS. PARKER: It has to do exactly with what you stated. That's why I thought you read it, because you quoted it perfectly.

CO-CHAIR ROSENZWEIG: Okay, fine.

MS. PARKER: High use of costs.

CO-CHAIR ROSENZWEIG: No, no, but I mean, but I didn't think that -- they didn't regard them as outliers, as much as sort of a separate high cost item.

I don't think they thought of them as totally outliers, because there are so many of them, and diabetes diagnosis is so common.

So, I thought that they -- it was like, they couldn't compare them with the rest of the population, since they are really trying to -- this is a chronic care measure, okay.

CO-CHAIR CURTIS: Let me ask you then, so, to the -- I mean, it sounds like ideally, you would want to stratify it by new onset of diabetes, but you're not going to be able to have that.

So, if you look at these criteria, as a simple cardiologist, I need you guys to tell me, is the passing the sniff test? Is

this a reasonable set of decisions that they've made to identify population with --

MS. PARKER: In the management phase.

CO-CHAIR CURTIS: In the management phase --

MS. PARKER: -- of diabetes.

CO-CHAIR CURTIS: -- of diabetes?

MS. PARKER: That is their intent. I think they do that.

CO-CHAIR ROSENZWEIG: I think they actually have thought this through, and I -- I don't disagree with them on -- at least, with respect to this particular measure.

One would like to know about the data on costs, on patients who are newly diagnosed.

One thing I should clarify is that newly diagnosed patients with Type 2 diabetes are not the same as patients with new onset of diabetes.

The average patient with diabetes

is diagnosed four to five years after the onset of the disease.

So, there is a tremendous amount of undiagnosed diabetes out there. So, I would just -- I don't really object, myself, to their rationale for excluding these people.

MS. PARKER: Great.

CO-CHAIR ROSENZWEIG: There is a lot of diabetes education issues. There are a lot of counseling issues. The frequency of visits is much more frequent in the first six months after diagnosis of diabetes for most people.

MS. PARKER: So, just to summarize, it sounds like the inclusion criteria that they have proposed within the document is appropriate.

CO-CHAIR ROSENZWEIG: Except for, they've left out medications --

MS. PARKER: Well, they get to medications, later. So, we'll --

CO-CHAIR CURTIS: No, but they

used specific inclusion medication.

CO-CHAIR ROSENZWEIG: There are certain medications that could -- that should be included for --

MS. PARKER: Oh, you're saying the non-insulin injectables, perhaps, because they just say one oral hypoglycemic or one insulin --

CO-CHAIR ROSENZWEIG: Yes, I think they may have left them out because they are not usually first-line agents.

MS. PARKER: Sure.

CO-CHAIR ROSENZWEIG: They're usually second- or third-line agents.

MS. PARKER: But it is the maintenance phase of the --

CO-CHAIR ROSENZWEIG: But we're in the maintenance phase, anyway.

MS. PARKER: -- for diabetes, so, yes.

CO-CHAIR ROSENZWEIG: Yes.

MS. PARKER: So, that needs to be

clarified within the inclusions.

CO-CHAIR ROSENZWEIG: Right, yes.

MS. PARKER: Okay, perfect. So, now, we go to the exclusion criteria.

No surprises here: PCOS, gestational diabetes or steroid-induced diabetes, cancer, ESRD, renal failure, HIV/AIDS and organ transplant.

Now, on the flip side of the newly diagnosed patients who have a lot of costs, they've excluded kind of the other end of patients with diabetes, as far as renal failure, end stage renal disease, and my recollection of diabetic nephropathy is that it's -- diabetes is the leading cause of liver failure, liver issues, in general.

And so, it's interesting that, as you'll see later on, they include other conditions that are kind of linked with the microvascular conditions, retinopathy, neuropathy, but they leave out nephropathy.

So, that was just kind of

something that stood out to me, and again, in the context of yesterday's conversation, where, you know, in diabetes, why was end stage renal failure left out? Why was organ transplant left out, when they have pancreas and kidney transplant?

So, just opening that up to the panel, to discuss, and I don't remember where we landed yesterday, with kind of keeping those out of the proposed -- I think it was the NCQA measure.

So, it may help to know kind of where we landed there to guide where we should land here.

DR. HWONG: I think in general, for that, it was really just about sort of the high costs, in terms of the ESRD population --

MS. PARKER: Right.

DR. HWONG: -- in terms of those cost outliers. I think what is interesting here is that they also include this category called renal failure.

MS. PARKER: Right.

DR. HWONG: So, you know, maybe, you know, Brenda, you know, looking at this, maybe the measure developer can answer, but what -- how are you defining renal failure?

Like, what chronic disease stage is included in that exclusion criteria?

MS. PARKER: And I don't know off the top of my head. I don't know if they -- I don't know --

DR. HWONG: Yes, maybe it's been --

MS. PARKER: -- the codes well enough, to know what the codes mean.

DR. HWONG: Sure, if the measure developer maybe could help us out with that.

DR. WEISS: Yes, it's three specific ICD9 codes. You know what? I'm going to have to look them up to be able to tell you exactly what they are.

DR. HWONG: Okay.

MS. PARKER: So, I think that is -

-

DR. WEISS: It's 585.2, 585.3 and 585.4. I mean, I don't know if that helps anybody, but I can get you a --

DR. WEINTRAUB: Those aren't ICD9s.

DR. HWONG: Yes, 585 point what?

DR. WEISS: It's 585.2, 585.3 and --

CO-CHAIR CURTIS: Five-eighty-five-point-four is chronic kidney disease stage four.

MS. PARKER: Okay, so, that makes sense, as far as being on the more severe spectrum.

CO-CHAIR CURTIS: But then why wouldn't end stage diabetic retinopathy or blindness be excluded?

MS. PARKER: And that was --

CO-CHAIR CURTIS: So, the ESRD, I understand, and we accepted it --

DR. HWONG: Yes.

CO-CHAIR CURTIS: -- for the NCQA measure, because it's such a high cost area.

This is more trying to homogenize the clinical severity of diabetes with fairly arbitrary thresholds.

MS. PARKER: Well, yes, and then --

DR. HWONG: Yes, this gets to --

MS. PARKER: -- kind of, somewhat of a normal -- I mean, the very first stage of CKD, I believe, from recollection, is -- I mean, it's not a terribly low creatinine clearance.

So, it's something that I would assume would be relatively common, perhaps.

DR. HWONG: Right.

MS. PARKER: More so then it warrants being excluded.

So, I think -- and that was my concern again, not including nephropathy, when neuropathy, retinopathy are included and nephropathy would seem to be one of the most

important aspects of diabetes, and maybe that's a stratification or sub-group or something.

CO-CHAIR ROSENZWEIG: Well, they're including nephropathy, they're just not including the ESRD.

MS. PARKER: Well, they actually don't mention diabetic nephropathy specifically as they do -- and this is by words only -- as they do in the identification of related services, they mention diabetes, poly-neuropathy, diabetic retinopathy and diabetic cataract.

CO-CHAIR ROSENZWEIG: They don't mention any aspects of -- where are you?

MS. PARKER: I'm sorry, I am on page 11, the fourth paragraph from the bottom, the first and second lines.

CO-CHAIR CURTIS: So, the identification of diabetic related services?

MS. PARKER: Yes, exactly.

DR. HWONG: I sort of mentioned

this point before, and again, sort of it causes -- you know, me to be a little bit concerned, again, sort of adding this renal failure category and sort of opens this question about what other types of complications would you want to add.

But you start to sort of whittle down this population a bit, especially if one of the codes, I think I just put in there, I can be wrong, it's -- you know, I'm looking ICD9 look up on the internet.

But if it's like CKD stage three, as well, I mean, you're going to -- so, you're sort of truncating this group and again, I understand it's in the effort of being homogenous and that's all good, I love things that are comparable.

But I just -- you know, I sort of wonder, in the end, so, you're taking away, again, these newly diagnosed folks.

You're taking away, you know, on the other end of the spectrum, you know, some

individuals, not only sort of extreme costs, but also, you know, individuals with some evidence, you know, of moderate chronic kidney disease and it just starts to get -- starts to feel like a much sort of smaller, narrower group.

MS. PARKER: Sure.

CO-CHAIR ROSENZWEIG: Well, later on, they do mention diabetes and renal complications. I'm looking at page 12, which is what, I guess, you were referring to.

CO-CHAIR CURTIS: I think we're referring to the exclusion of these patients, and it -- and 585.2 is mild.

So, you know, they are excluding the gamut of patients with renal insufficiency, at least as diagnosed by these specific ICD9 codes, and I agree, it's sort of -- you know, why not exclude heart failure patients? You know, it's sort of --

CO-CHAIR ROSENZWEIG: Well, this is --

CO-CHAIR CURTIS: How

homogenous does the population have to --

CO-CHAIR ROSENZWEIG: -- there is end stage renal disease, and then they have including dialysis, but where are they excluding --

CO-CHAIR CURTIS: Well, it's --

CO-CHAIR ROSENZWEIG: Where are they excluding --

MS. PARKER: Renal failure is directly below end stage renal disease and --

DR. HWONG: On page 16.

MS. PARKER: -- and dialysis.

DR. HWONG: It's one of the bulleted --

MS. PARKER: On page 11.

DR. HWONG: I'm sorry, I see it in another area, too.

MS. PARKER: Yes, and I think from my research experience, you know, renal disease is an important -- I mean, we always look at it in terms of the sub-groups and the different classifications of CKD and that was

important.

It may be because my research was sort of looking at that sub-group individually, but it's always been something important.

CO-CHAIR ROSENZWEIG: Yes, I would say that chronic kidney disease certainly shouldn't be excluded.

I think one of the issues that does come up with end stage renal disease, and dialysis is that those patients tend to go to a different pool --

MS. PARKER: Sure.

CO-CHAIR ROSENZWEIG: -- insurance pool. They end up in Medicare.

MS. PARKER: Right, right.

CO-CHAIR ROSENZWEIG: And so, they are pulled out of the -- the costs are actually pulled out of -- if this is designed mostly, let's say, to apply to commercial insurance, they're in a different category.

So, but I'm not justifying that,

necessarily, but the other basic issue is that the two biggest cost drivers for diabetes are chronic renal disease and -- but even -- and cardiovascular disease.

So, we're not excluding cardiovascular disease.

MS. PARKER: We don't. They actually mention hypertension and hyperlipidemia, as important areas within this population to address.

CO-CHAIR CURTIS: I guess, we would ask the developer, then, is to sort of provide a more complete justification of this decision as to what -- and specifically, you know, I think it might affect kind of how our take on the measure is, so it's fairly important.

CO-CHAIR ROSENZWEIG: And later, in their accounting of the costs, they do include diabetes, renal-related codes.

So, if they're excluding them up on top, why are they including them later on?

MS. PARKER: Yes, and my next notes, actually, just confirming the accuracy, consistency of the codes that they recommend.

MS. CLARK: Yes, I actually had a question about all these exclusions, anyway, because if you're risk-adjusting the costs anyway using the HCC scoring, then do you need to exclude them?

You know, they are already addressed in the --

CO-CHAIR CURTIS: Well, I think we've already sort of accepted the template of the NCQA saying --

MS. CLARK: Yes, I know, I'm just --

CO-CHAIR CURTIS: -- you know, like saying these are reasonable high-cost areas that can't be really adjusted for, on the basis of the HCC's or other risk-adjustment methodology.

So, I think where -- there needs to be specific rationale, is to where they

diverge from that preset cohort, which may not be complete, but at least explain why you're adding onto that.

CO-CHAIR ROSENZWEIG: Yes, I mean, if they want to exclude ESRD and dialysis, that's one thing, but they cannot really exclude the large proportion of these patients that have microalbuminuria or pre-ESRD -- up to pre-ESRD.

MS. PARKER: I agree, and just to the earlier point of making sure that the list of medications that are to be evaluated are complete, and that's something that we have identified as a deficiency within the medications listed.

One question I did have for NQF is, with all of these measures, how are they updated when new medication classes or products are -- or codes, even, I mean, how are these kind of maintained to update?

CO-CHAIR CURTIS: Let me just --

DR. HWONG: Yes.

CO-CHAIR CURTIS: I think you would -- that would fall under measure maintenance in the three year reviews.

If, in the interim, however, events occur, such as the release of a whole new class or some other key thing, which changed the measure definition, there are -- we can do it on a more frequent than three year basis.

MS. PARKER: Okay, perfect.

MS. TURBYVILLE: Right, and we actually just started a continuous annual update process, as well --

MS. PARKER: Okay.

MS. TURBYVILLE: -- in addition to the maintenance review.

MS. PARKER: Okay, sorry, I've had that question all along. I just waited until my turn to speak.

So, moving along, I won't belabor, again, I just think that all of the codes need to be confirmed to be accurate and consistent,

in the inclusion/exclusion and the codes they have listed, and --

CO-CHAIR CURTIS: Did you feel like any class of outcomes or costs were missing, in this current data set, any, I guess broad costs, is --

MS. PARKER: Nothing that jumped out at me, but again, I don't manage patients with diabetes often, so, there may be some nuances that I did not capture, based on my unfamiliarity of management of patients with diabetes.

CO-CHAIR ROSENZWEIG: One thing that came up, that is a high cost item is bariatric surgery.

MS. PARKER: And is that in here or not in here?

CO-CHAIR ROSENZWEIG: It's now indicated -- it should -- it's not in there, not that I could find.

MS. PARKER: Okay.

CO-CHAIR ROSENZWEIG: I don't -- I

checked the codes, to the best of my ability.

MS. PARKER: Okay.

CO-CHAIR ROSENZWEIG: But it is now approved for use in patients with diabetes and a lower obesity category, than those without diabetes.

MS. PARKER: Okay, perfect, thank you. One question I had, with regards to kind of our concern with the medications, that was kind of where I'm comfortable, in those instances, is that hypertension, hyperlipidemia are called out as being important, in terms of identifying patients with diabetes and their co-morbidities. Nephropathy was not.

However, and you could argue that this is in terms of hypertension, you know, ACEs and ARBs are included, that prevents nephropathy, but while neuropathy was called out as an important -- and this is -- it gets confusing, neuropathy was called out as important, I don't really see any medications

that are specific to neuropathy.

There aren't that many that are approved specifically for diabetic neuropathy. There are some that are used off-label for it, and are very effective, however, the one that came to mind was Lyrica, I believe, is the specific product drug that is approved for diabetic neuropathy.

So, just kind of, again, a disconnect between what we're saying is important in the medications that were indicated, that there seems to be kind of a difference that jumped out at me.

I don't know if that is important or it's just something that the --

MS. TURBYVILLE: Do you want to ask the measure developer?

MS. PARKER: Sure, that's --

MS. TURBYVILLE: If that was intentional.

MS. PARKER: Yes, measure developer, was that intentional?

DR. WEISS: To exclude those, right? Our expert panel, our clinical work group, identified the medications that they were interested in including, and those drugs were not on their list.

So, there is not an intentional exclusion. They didn't come out and say, "We don't want to include these drugs." They did not show up on our frequently used medication list, that were not grouping to our episodes.

MS. PARKER: And perhaps in the interest of time, this is just another statement, or another scenario that underscores the difficulties of identifying diabetes related, rather than just taking all of the resources, as was proposed in a previous measure.

So, it's a casualty of the method selected, perhaps.

CO-CHAIR ROSENZWEIG: There are a whole variety of medications for, yes, treatment of painful diabetic neuropathy, that

are not included here.

MS. PARKER: So, that just may be a deficiency that the Steering Committee will need to decide, if it's acceptable or that we would ask that they go back and address, I think.

CO-CHAIR CURTIS: But as the expert reviewer here in this --

MS. PARKER: I think they need to be --

CO-CHAIR CURTIS: -- do you think it's a --

MS. PARKER: I think if they're going to say, "These are the diabetes specific medications," then you need to make sure that you have every diabetes specific medication on the list, in my opinion.

CO-CHAIR CURTIS: Or rationale for the exclusion.

MS. PARKER: Exactly.

CO-CHAIR CURTIS: Okay.

CO-CHAIR ROSENZWEIG: And there

is, you know, at least one medication for treatment of peripheral vascular disease.

MS. PARKER: Yes, and again, that is diabetes related, so, I think that this warrants a revisit.

Then lastly, and I know it seems like it's hard to say lastly, within this section, is the stratification.

So, this gets back to, you know, are there sub-groups that should be stratified or -- because there are no stratification measures proposed, no stratification at all.

At the very least, with some of the others, we've seen stratification based on populations, disparities, you know, some sort of stratification, and maybe this again --

DR. WEINTRAUB: You don't need to.

MS. PARKER: -- this goes back to the need for perhaps, a clarifying definition for stratification, because there seems to be some confusion on is stratification something that you do in the very beginning or

essentially, you have this full group and you stratify on one variable that could impact outcomes, or does stratification refer to the sub-grouping afterwards, where you report the information, based on different sub-groups that have been identified?

DR. WEINTRAUB: Well, you know, I think --

MS. PARKER: And it's used both ways, unfortunately, in the public domain, honestly.

DR. WEINTRAUB: Clarification of why you would want to stratify is needed.

Again, the real reason for -- I guess you could come up with two reasons for stratify.

One reason would be that is aesthetic. So, if you consider Type 1 and Type 2 diabetes, you might not want to have them together --

MS. PARKER: Exactly.

DR. WEINTRAUB: -- or ST segment

elevation or non-ST segment elevation, but that is aesthetic.

The other reason is analytic, that you -- in developing your model, you may want to stick with main effect models, and not have interactions, specific interactions are very confusing to people.

MS. PARKER: Absolutely.

DR. WEINTRAUB: And so, if you have interactions, you can still put them all in one pot, and deal with it that way, and mathematically, it will work out.

MS. PARKER: Right.

DR. WEINTRAUB: But you may choose not to do that. Those are the only reasons for stratify.

Looking at sub-groups, on the other hand, is something that's perfectly fine to do.

So, for instance, in dealing with patients with diabetes, you may want to look at the sub-group with peripheral vascular

disease, or look at the sub-group with heart failure or what have you, or want to look at them all together.

In all of these measures, this kind of analytic approach, none of them -- in none of the ones we've discussed here, has this been laid out as a kind of analytic strategy.

MS. PARKER: Right, I would agree.

CO-CHAIR CURTIS: But for the specific measure, they're not specifying any --

MS. PARKER: They said nothing.

CO-CHAIR CURTIS: -- stratification necessary, based on the -- their efforts upstream, to make this --

MS. PARKER: Right.

CO-CHAIR CURTIS: -- a more homogenous population.

MS. PARKER: Right.

CO-CHAIR CURTIS: Okay.

MS. PARKER: Which again, is my --

some of my confusion is just the intense case they made for the disparities in race.

So, but again, I think we need to clarify, because I think it's been confusing, and this has come up, is the intent of the question, stratification from an analytic perspective, or is it a sub-group from reporting.

CO-CHAIR CURTIS: Not to interrupt, but I think what we've heard and I think what we have to take back to the Steering group level, and maybe even higher --

MS. PARKER: Sure.

CO-CHAIR CURTIS: -- within NQF is, is it important to address --

MS. PARKER: Right.

CO-CHAIR CURTIS: -- disparities within resource use --

MS. PARKER: Right.

CO-CHAIR CURTIS: -- and you could see it as being part of something that exacerbates disparities, but is it something

that you need to report separately? I'm not sure.

I think it's more important for process or outcomes measures, but we'll take that up to the next level.

MS. PARKER: Yes, and not a deal breaker, from my perspective. Again, I think it's very ambiguous, as to what the definition is and the intent is, and I think it would be helpful, though, for measure developers in general to kind of weigh in as to how they think that this would be reported beyond some broad provider summary category.

You know, is it -- and that would demonstrate a true, I think, understanding of the disease state, and that they did use kind of key opinion leaders or experts in the field, to understand how this would be reported, or how the user may find it useful to look at this, just beyond the peer group, but maybe also within the disparities or the sub-groups.

That's my opinion. I don't think it makes or breaks this --

MS. CLARK: I like the idea that the NCQA had, of providing information on percent of patients or the number of procedures along with the cost information.

So, you know, for example, if you're going to create a report, looking at the different cost categories, why not also provide a report looking at the distribution of patients within -- you know, within these certain categories that you're talking about, at least?

I mean, that might be helpful, additional information.

CO-CHAIR ROSENZWEIG: It's 2a2 and 2b2, is that correct?

MS. PARKER: Yes, I believe so.

CO-CHAIR ROSENZWEIG: Okay.

CO-CHAIR CURTIS: And 2b1.

MS. PARKER: Did we want to let the measure developer comment, or ask anyone

else for any questions, before we move on to voting?

CO-CHAIR CURTIS: I propose we just move to vote on these two. I think you did a nice job of leading us through the differences, and I think we have a good understanding.

MS. PARKER: Okay, great.

CO-CHAIR ROSENZWEIG: Okay, so, 2b1 is the measure specifications are consistent with -- MS. TURBYVILLE: Two-a1.

MS. PARKER: Two-a1.

CO-CHAIR ROSENZWEIG: I'm sorry.

MS. TURBYVILLE: That's okay.

CO-CHAIR ROSENZWEIG: Two-a1 is the measure is well defined and precisely specified, so it can be implemented consistently within and across organizations and allow for comparability.

MS. PARKER: Can we see the comparison on the screen, of the --

MS. TURBYVILLE: Sure.

MS. PARKER: I know it's a different disease stage, but within the context of the disease stage.

MS. TURBYVILLE: So?

CO-CHAIR CURTIS: Do you want to bring that up over -- what you recommend for this?

MS. PARKER: Sure, I was actually on the sense of moderate to low, just based on what we've discussed previously, with regards to the similarities.

But given the concerns within just the clinical framework itself and the construct of the measure, I would probably vote low on this, at this time, because I think there is some room for improvement.

CO-CHAIR CURTIS: So, maybe I'm --

MS. PARKER: No, go ahead.

CO-CHAIR CURTIS: I think low is a specific threshold that sort of -- it's a -- well, it's a --

MS. PARKER: I can vote moderate.

CO-CHAIR CURTIS: It's a barrier to moving forward and I mean, I just -- this is my opinion.

MS. PARKER: Sure.

CO-CHAIR CURTIS: I feel like, we've identified ways that they could improve it or refine the measure.

MS. PARKER: Sure.

CO-CHAIR CURTIS: But I don't think we've found anything that we could characterize as a fatal flaw.

MS. PARKER: As a critical flaw, sure.

CO-CHAIR CURTIS: Again, my opinion, but --

MS. PARKER: And so -- go ahead.

DR. MARWICK: The chronic kidney disease issue is a significant piece.

CO-CHAIR CURTIS: I think it's significant, not ignorable, but I don't know if it's --

MS. PARKER: But you think that

that's very easier --

CO-CHAIR CURTIS: That's fixable.

MS. PARKER: That is easily --

CO-CHAIR CURTIS: It's easily fixable.

MS. PARKER: Okay.

CO-CHAIR CURTIS: Or they could just clarify --

MS. PARKER: Sure, and that was what I was thinking, was that the kidney issue was a serious concern, but if it's -- if the addressability of it means that we could vote moderate, because it is something that's easily fixed, then I think that's --

CO-CHAIR CURTIS: We could respond to it and --

MS. PARKER: That's fair.

CO-CHAIR CURTIS: Yes.

MS. PARKER: That is a fair assessment.

CO-CHAIR ROSENZWEIG: So, we're not -- we're not really commenting on errors

in their definition? It's whether it can be defined? Is that it? I'm a little confused here, because --

MS. PARKER: Well, they haven't defined it, really.

CO-CHAIR ROSENZWEIG: Really, we've identified a whole variety of things that --

CO-CHAIR CURTIS: I mean, I'm not saying how we should -- how you should vote. I just think --

CO-CHAIR ROSENZWEIG: Type 1 versus Type 2, those kinds of issues are --

MS. TURBYVILLE: That is validity, I think. This is about -- so, just to -- this is really about how precisely the specification is written, and then you'll get into whether or not they included the right codes more and the validity and -- right, so --

MS. PARKER: But the specification of diabetes is --

CO-CHAIR CURTIS: It's along the sort of --

MS. PARKER: It's pretty broad.

MS. TURBYVILLE: Okay, okay.

CO-CHAIR CURTIS: We're including that, as in part of the specifications.

MS. PARKER: Right.

CO-CHAIR CURTIS: It's like, not just how well -- how precisely specified it is, but how accurately that reflects the population, the target population.

CO-CHAIR ROSENZWEIG: But we are --

CO-CHAIR CURTIS: So, we've identified it --

CO-CHAIR ROSENZWEIG: We're identifying the -- we were telling them to change the title of this, the Type 2 diabetes.

MS. TURBYVILLE: Right, but you're going to want to make sure that that comes up again, then, in 2b1, which is about, is it consistent with the evidence, and 2a2 is

really about how it's written and can it be implemented consistently?

CO-CHAIR CURTIS: Okay.

MS. PARKER: Two-a1, you're saying is implemented?

MS. TURBYVILLE: Yes, is it written clearly enough, which I think you guys have identified themes across measures, to be implemented consistently, and 2b1 definitely is the place where you're saying, you know, we're talking about people with diabetes, some are being carved out, some are -- you know, all of the conversations that you have had.

MS. PARKER: And that makes sense within the context of reliability, as we discussed yesterday, that it has been tested and it can do it, with its testing. So, that makes sense, in the reliability, in looking at it in the broad sense.

MS. TURBYVILLE: Not that it doesn't influence.

MS. PARKER: Sure.

MS. TURBYVILLE: Clearly, what you're looking and the precision, I completely agree. I just want to make sure you also think about that in 2b1.

MS. PARKER: Okay.

CO-CHAIR ROSENZWEIG: All right, so, for 2a1, what is your recommendation?

MS. PARKER: After the very thorough explanation by NQF, I think I would still go with moderate, because I still think that there is some room for improvement.

CO-CHAIR CURTIS: Let's go ahead and vote.

CO-CHAIR ROSENZWEIG: Let's vote. Six moderate and two low.

All right, then 2b1, is that correct, is the next one?

MS. TURBYVILLE: Yes.

CO-CHAIR ROSENZWEIG: The measure specifications are consistent with the evidence presented, and support the focus of measurement under criterion 1b, and the

measure is specified to capture the most inclusive target population indicated by the evidence, and exclusions are supported by the evidence.

MS. PARKER: So, here, I think it's more of the issue of distinguishing between Type 1 and Type 2, as well as the exclusion of renal failure.

I think those are our two major issues. Can they be easily fixed? I don't know, that would be something that the measure developer would have to weigh in on, but at this time, I think moderate to low is going to, for me, go to low.

Again, that's not a judgment or an indictment on anyone else, that they need to do the same.

CO-CHAIR CURTIS: Okay, let's vote. I'm becoming more like a surgeon.

CO-CHAIR ROSENZWEIG: All right, two moderate and six low.

MS. PARKER: So, going back to

2a2, reliability, again, this gets at not if what they necessarily have is right, but if what they have currently was tested sufficiently, to demonstrate repeatable results, and they use the Market Scan database, which is a large database, that has lots of patients with diabetes.

So, they had a large population to work with, and for me, I think that based on the consistency of the results, that they were -- that they presented in their -- even, you know, removing some of the pieces and modifying the measure some, produced consistent results.

I was okay with this. I don't know if within some of the other measures we've looked at, their slides, if there would be any need to go through those with the panel, given that it's kind of the consistent -- the same reports, the same slides, the same data presented, just in the context of diabetes.

Is there anything the measure developer would care to add to the discussion?

DR. WEISS: The only additional piece of information on testing is that we also tested our diabetes measure in a large data set in Wisconsin, that we have data -- have found that similar performance within that data set as we feel was in the Market Scan data.

MS. PARKER: That was in a Wisconsin specific database? Was my reading of that correct, that it's Wisconsin, which I don't think is a largely populated state.

So, I would just -- I didn't think that was as strong as -- and I could be wrong. I'm not from Wisconsin, I don't claim to know much about Wisconsin, other than Brett Favre.

So, you know, I'm definitely limited. So, I didn't just see that as kind of an overwhelmingly credible database. Maybe it's just my lack of knowledge.

CO-CHAIR CURTIS: No, it's 3.4

residents, million residents, 207 million claims against -- I mean, it's not ignorable.

DR. WEINTRAUB: That's not bad.

MS. PARKER: So, there are a lot of cheeseheads, clearly.

DR. WEINTRAUB: This has been tested in a good size program.

COURT REPORTER: Use your microphone.

DR. WEINTRAUB: This has been tested in several good size cohorts.

MS. PARKER: So, if there are no further comments, we'll --

DR. WEINTRAUB: Well, so, they haven't the --

MS. PARKER: Microphone.

DR. WEINTRAUB: Yes, they have the same kind of problem with related/non-related services that we've seen before.

If you go to slide eight, it will -- you will see for, right at the top, routine gynecological examine, they have related/non-

related services, routine medical exam, related/non-related, chest pain, related/non-related.

And I think that they have some problems here, in making sense out of that, same kind of thing that we saw before.

CO-CHAIR CURTIS: And I think that's, you know, not that we're parking lotting it, but that it's consistent with the other ABMS --

MS. PARKER: Exactly.

CO-CHAIR CURTIS: -- or REF measures that we've addressed.

But specifically, with regard to 2a2, results are repeatable. In fact, this is actually the highest test, where they've looked at kind of --

MS. PARKER: Yes, exactly.

CO-CHAIR CURTIS: -- looked at comparable data sets.

MS. PARKER: Exactly.

CO-CHAIR CURTIS: Not having to

redefine the costs, based on --

MS. PARKER: And I looked at this, after this morning's discussion, just to confirm that yes, they didn't have to change anything, but they also weren't using necessarily different -- commercial and Medicare populations are different.

So, it kind of makes sense that they would to perhaps, change methodology. It doesn't make it easy, but intuitively, I get it.

Here, there really was no obvious difference in the databases that would warrant potential changing.

So, no, I think they did a great job of testing the reliability of the measure.

DR. WEINTRAUB: Yes, so, the other thing that makes that -- that will make this work is the distribution is pretty reasonable, if you go to slide 17.

MS. PARKER: Sure.

DR. WEINTRAUB: But drug charges

is,
by far and away, the number one cost.

But E&M, not a durable medical equipment -- and I think the other medical equipment was a little more of a problem, but they're really not too bad.

CO-CHAIR ROSENZWEIG: Which slide are you referring to?

DR. WEINTRAUB: Slide 17. The other things we're very concerned about in developing the measure was thinking about in-patient facility charges, but the 99 percentile is still zero dollars. Not a lot of -- not a lot of hospitalizations in these folks.

MS. PARKER: And I believe that that comment was made earlier, that this is mainly an out-patient sort of disease, and that that wouldn't be terribly high, although requiring an in-patient -- no, an in-patient wasn't required for this, my apologies.

CO-CHAIR ROSENZWEIG: No.

MS. PARKER: It was just, it could be counted as one of the resource use requirements.

So, I think we're okay, with 2 --

CO-CHAIR ROSENZWEIG: Actually, I'm surprised that there was a zero in-patient facility charge.

MS. PARKER: Well, the mean was --

CO-CHAIR ROSENZWEIG: Somebody was --

MS. PARKER: -- terribly low.

CO-CHAIR CURTIS: Well, it doesn't say no, but it means that --

MS. PARKER: The mean is $215, so, you don't have to --

DR. WEINTRAUB: This is 95 percent comparable, so the --

MS. PARKER: And this normally rounds down and up.

DR. WEINTRAUB: That means it's still a couple -- a couple of percent of the people that are hospitalized, and that is not

unreasonable.

CO-CHAIR CURTIS: But it would be nice if we could see the range on that, to see if that improves, or what percent were actually hospitalized, might be useful feedback.

MS. PARKER: Well, and that might be, again, back to the point earlier, by looking at frequencies, as well as costs, and how NCQA did it, as well. So --

CO-CHAIR ROSENZWEIG: So, this is very different from the Medicare population?

MS. PARKER: Absolutely.

CO-CHAIR ROSENZWEIG: Where the big cost drivers are actually hospitalizations.

MS. PARKER: Okay, so, do we vote now on 2a2, or do we go to 2b2 and vote on those, together?

CO-CHAIR CURTIS: Let's keep going.

MS. PARKER: Keep going, okay,

great. So, 2b2, I think this validity gets back to 2b1, and the concerns that we have there, with it being that the data elements are -- there are some significant room for improvement with the clarification of the data elements with 2b2.

CO-CHAIR CURTIS: But what about -- so, if you look at 17, to me, at least, there is some face validity to that, as to that they are clinically meaningful and important differences in cost?

MS. PARKER: And granted, I actually rated that as moderate, because of the concerns that I had within the data elements being correct. But yes, it does seem to be valid.

CO-CHAIR CURTIS: Okay.

MS. PARKER: Any other comments or questions?

Okay, 2b3, exclusions are supported by clinical evidence, measure specifications for scoring include computing

exclusions, so that the effect on the measure is transparent, and patient preference.

I don't think that that necessarily applies here. So, if -- and I'll kind of walk everyone through that, if -- quickly, if that's desirable.

But essentially, I just noted that they have not sufficiently -- I mean, they've tested it in the cohorts, in the databases that they have, but there were still some concerns with the exclusion criteria of renal disease and that being impactful.

So, I still ranked that as moderate, being that they could improve that, and based on the previous discussion that improving the criteria would be a relatively easy thing that they should address.

And then 2b4, if there are no questions, moving along here, risk adjustment method.

It seems to be that the risk adjustment methodology is widely accepted

among all the measures, no difference here in my opinion, and getting back to the stratification issue, I think that's still something that has been put in the parking lot, as something that will be addressed, as to if this is really important.

So, here, I would rank this still as moderate. Oh, actually, I think it would probably change that to high, given that we have agreed on HCC, and with the caveat that the stratification issues are still something to be determined by the Steering Committee.

DR. WEINTRAUB: So, we have some kinds of modeling issues here, that we don't see in our -- unless I'm missing it, we don't see that R-squared here, not only that, they could the R-squared in the validation population, which would really be nice, and they don't have calibration here.

MS. PARKER: And that's something that they've been requested to provide, is that correct?

DR. WEINTRAUB: Yes.

MS. TURBYVILLE: For all the measures.

DR. WEINTRAUB: For all the measures.

MS. PARKER: For all the measures, okay.

DR. WEINTRAUB: But here, I'm going to say that they can do a -- they've got the second cohort, so, they can do proper validation of the models --

MS. PARKER: Sure.

DR. WEINTRAUB: -- to be -- step up, yes.

DR. MARWICK: Once they're familiar with the risk adjustment process -- is heart failure a part of that, do you know?

MS. PARKER: I'm not sure if it is included.

CO-CHAIR CURTIS: Yes, it is.

DR. MARWICK: It is?

CO-CHAIR CURTIS: It's one of the

HCCs.

MS. PARKER: Okay. Okay, 2b5, here, I interpreted this a little differently than I think most people have, in that I looked at the type of score you're using, as well as the interpretation of the score, and it looks like, you know, based on what I've read is that the score they're using is actually the observed to expected ratio, which has been accepted with all of the other measures, that have been proposed.

So, in my opinion, I thought that, you know, based on the consensus of the panel of previous measures, that it was an acceptable way to identify these, and it did provide a meaningful comparison among the groups, that they have provided in their data, whether it's region, provider, state, et cetera.

So, unless I missed something significant, I thought that was completely appropriate and that they did valid that and

make sure that that does provide meaningful information.

CO-CHAIR ROSENZWEIG: Any comments? Okay.

MS. PARKER: Two-b6 doesn't really apply, and 2c, I think here, it is going to be a similar vote, as to the other proposed measures by the measure developer.

CO-CHAIR CURTIS: Right, so, we'll take that up to the Steering Committee.

MS. PARKER: Exactly.

CO-CHAIR CURTIS: And vote on it, yes.

CO-CHAIR ROSENZWEIG: Okay, let's do the voting on these measures, on these components. I guess we start with 2b2?

MS. PARKER: Two-a2.

CO-CHAIR ROSENZWEIG: Two-a2, I'm sorry, I keep on forgetting.

MS. TURBYVILLE: Two-a2.

CO-CHAIR ROSENZWEIG: All right, so, 2a2 is reliability testing demonstrates

the results were reproducible, producing the same results in a high proportion of time, when assessed in the same population, in the same time period, and that the measure score is precise.

MS. PARKER: Yes, and they did demonstrate that it is reliable, the way that it is, using the two databases.

CO-CHAIR ROSENZWEIG: Yes.

MS. PARKER: So, I voted high for that one.

CO-CHAIR ROSENZWEIG: Especially if you live in Wisconsin.

MS. PARKER: That's right.

CO-CHAIR ROSENZWEIG: Okay, now, 2b?

MS. TURBYVILLE: Two.

CO-CHAIR ROSENZWEIG: Okay, validity testing demonstrates that the measure data elements are correct and the measure score correctly reflects the costs of care for resources provided, adequately distinguishing

higher and lower cost resource use.

MS. PARKER: So, in their results, they did show that it was valid, the way that it was tested, but there are some existing concerns with some of the data elements in the definitions and inclusion and exclusion.

So, based on that, I would rank it as moderate.

DR. WEINTRAUB: Now, you know, we don't see a formal calibration.

MS. PARKER: Right.

CO-CHAIR ROSENZWEIG: All right, so, they're all moderate, okay.

Okay, 2b3, that's where we are on the next one?

MS. TURBYVILLE: Yes.

CO-CHAIR ROSENZWEIG: Exclusions supported by the clinical evidence, otherwise, they are supported by evidence that sufficient frequency of occurrence of the results are distorted with the exclusion.

MS. PARKER: I think the exclusion

criteria makes sense, for the most part, based on the latter part of that, and that they have some of the other ESRD's, high cost.

However, the inclusion of renal failure as an exclusion criteria did raise some concern with the panel.

So, I will still vote them -- or rank this as moderate, noting that there is room for improvement.

CO-CHAIR ROSENZWEIG: One high and seven moderate.

DR. HWONG: That's me, I'm sorry, I miscounted. Could I hear it back?

MS. TURBYVILLE: I agree.

CO-CHAIR ROSENZWEIG: Sorry, which one?

DR. HWONG: I forget which one I was, by accident.

MS. TURBYVILLE: Okay, you guys ready? Go ahead.

DR. HWONG: Sorry about that.

CO-CHAIR ROSENZWEIG: That's

interesting. Brenda, I was going to say that you were doing as well as Kim Jong-il, but I don't think so.

All right, okay, so, we're up to 2b?

MS. TURBYVILLE: Five.

MS. PARKER: Four.

CO-CHAIR ROSENZWEIG: Four? Four, the risk adjustment strategy, 2b4.

Evidence based risk adjustment strategy is specified and is based on patient clinical factors that influence the measured outcome, but not factors related to disparities in care.

MS. PARKER: And so, again, just to reiterate, kind of our general consensus on the HCC being an accepted risk stratification, or adjustment method, I think this would be high, except for the fact that I still have some concerns with stratification and not understanding it completely, and I understand it will be parking lotted.

So, I still think based on what comes from the Steering Committee on that, there may be room for improvement. So, I would go with moderate, on this one.

So, there were three high and five moderate, and we will go to 2b5.

CO-CHAIR ROSENZWEIG: Yes, and this is the data analysis demonstrates that the scoring and the method -- the measure allows for identification of statistically significant and practically significant meaningful differences in performance.

MS. PARKER: And similar, and in my opinion, again, at least to 2b4, the OE ratio and its interpretation has seemed to be fairly accepted by the panel, and as presented in other measures, as it is here.

So, I personally ranked this as high, given its consistency with the other measure developers.

DR. WEINTRAUB: But they haven't done this.

MS. PARKER: They did. They provide the ratio and they provide the --

DR. WEINTRAUB: No, but the --

MS. PARKER: They do in the --

DR. WEINTRAUB: Say they do it. Say they do it, and then it comes --

MS. PARKER: No, if they look at --

DR. WEINTRAUB: But do they actually --

MS. PARKER: If you look in the slides, maybe -- and this would be great, because if I'm misunderstanding it, then that would obviously impact my interpretation, as well as my ranking.

However, if you look in -- let me get there, and measure developer, if I'm mis-representing you one way or the other, please let me know.

CO-CHAIR CURTIS: Slide 36 of the PDF.

MS. PARKER: Thank you. So, it

actually starts on 34, with -- they present their ratio, as they've calculated, by region, by state, by specialty, as they've done in all of the previous measures, and I thought that was meaningful, in looking at those values.

CO-CHAIR ROSENZWEIG: But wouldn't the issue of the fact that they're measuring at the provider level --

MS. PARKER: Well, we said earlier that that would --

CO-CHAIR ROSENZWEIG: -- and statistically significant issues, related to that part of this measure --

MS. PARKER: So, that was -- yes, that was something that NQF said that they would kind of look at, as making a blanket measure, as far as the interpretability and applicability of these, in the absence of statistical power.

Also, I believe that we recommend that perhaps, valid sample sizes could be calculated in the three year period, where

this is in use, and there is enough data to obtain that.

So, based on what we have here, I still think that the values are -- the way that it's scored, and the interpretation of the score, is meaningful, and it's something that's easily understood by most.

CO-CHAIR CURTIS: Also, just looking at this does make me a little bit more concerned about the peer group evaluations and the accuracy of the assessment of specialty, if they're -- you know, I don't know what the ratio of endocrinologist, internal medicine and family practice doctors is, but 5,000 seems low, and then if you have 5,000 endocrinologists and 3,000 cardiologists being captured by this measure, you do wonder if it's more of an issue than I had initially expected.

MS. PARKER: Well, and I think that still goes to kind of the sub-group and the -- not necessarily the score itself, but

as it is reported.

CO-CHAIR CURTIS: Yes.

MS. PARKER: So, I still rank this as high, because I think it makes complete sense.

CO-CHAIR CURTIS: No, they just raised that other issue.

MS. PARKER: Exactly.

CO-CHAIR ROSENZWEIG: Okay.

CO-CHAIR CURTIS: We're waiting for one response.

MS. TURBYVILLE: One more.

DR. WEINTRAUB: Everyone, hit your button six times.

MS. TURBYVILLE: There you go.

MS. PARKER: So, there were six high and two moderate, and I believe, correct me if I'm wrong, that the remaining measures, 2b6, 2c, all usability and feasibility fall along the same lines as before, is that correct, or am I --

CO-CHAIR CURTIS: I think we

should -- MS. TURBYVILLE: Yes.

CO-CHAIR CURTIS: -- yes, take the same approach as we took for the other.

MS. PARKER: Okay.

CO-CHAIR CURTIS: Either not applicable or we'll formalize the vote at the future date.

MS. PARKER: Okay, perfect, thank you.

CO-CHAIR ROSENZWEIG: Okay, thank you very much, Brenda.

MS. PARKER: Thank you.

MS. TURBYVILLE: Now, can we just get a statement from the TAP, about usability and feasibility for this measure, just so that we have it for --

CO-CHAIR CURTIS: So, just to formally state it, we would expect that, like the other ABMS area measures, that it's not been formally tested for usability, and we would likely have similar scores, but we'll formalize that in the future, similarly for

feasibility.

So, we have 11:45 a.m. So, I'm not sure if --

MS. TURBYVILLE: Yes, so, we'll open it up to public comment, now.

Operator, if you could open the line for any public input or questions at this time, we would really appreciate it.

OPERATOR: Certainly, that is *1, if you have a question or comment.

DR. LEE: This is Todd Lee. Can I make one comment that I think is relevant for all of our measures, that I've sort of learned through this process, over the last day and a half, while we're waiting for public comment, that I think we failed to do a good job communicating in our measure specifications.

The actual overall intent is to be able to provide actionable information with our measures and that's the reason that we focus on conditions that set the resource use.

So, that once you will provide, or

once a provider received a report that said maybe they're high or low on an O to E ratio, it would be able to go and use the data to find out why, and that's sort of the reports that under-lay the episode report at the O to E ratio for the position.

And so, we'd be able to certainly look, there is a lot -- we've got a lot of hospitalizations, so, you've got a lot of high cost medication use, or alternatively, if you're a low cost provider, then you're partnering that with a quality measure.

Now, this is -- you know, you compared our diabetes measure a lot to NCQA, and our group is completely different. Our measure is intended to say, "Look, here is the topic of diabetes," and what can you change possibly, if you're a high cost provider?

And I just think we did a good job -- or did a poor job, of communicating that to the panel, and I just wanted to be sure that we said that, as you consider our next couple

of measures.

CO-CHAIR CURTIS: I think that is fair. I mean, I do feel like we discussed that, certainly, and some of the previous measures have -- this is trying to be more actionable.

I think the concern has always been, you know, the specificity of the outcome and how complete it is.

So, but your point is well taken and acknowledged.

Are there any public comments?

OPERATOR: No, public comment at this time.

CO-CHAIR CURTIS: Okay, so, we're at a little bit of a crossroads. We have slightly less than four hours, before 3:30 p.m.

I'm not sure if people have to catch planes, or not, but we would like to respect that deadline, and we have at least two Ingenix measures that we would like to go

through, in that time frame.

So, what I would propose is sort of a natural break, but early lunch, and keep it as a very short lunch, and hope to be back by slightly after noon, 12:10 p.m., to get restarted, and that should give us a solid, almost three and a half hours to get through the two Ingenix measures.

(Whereupon, the above-entitled matter went off the record at 11:09 a.m. and resumed at 12:09 p.m.)

CO-CHAIR CURTIS: So, in the interest of maximizing our time together, is -- why don't we go ahead and get started on the Ingenix measure, on diabetes, that Jamie is going to take us through? Do you have the measure number?

CO-CHAIR ROSENZWEIG: Sure, this is measure number 1595, and the title of the measure is ETG-based diabetes resource use measure, and the measure steward is Ingenix, or how do you pronounce it? Is it Ingenix?

DR. LYNN: Ingenix.

CO-CHAIR ROSENZWEIG: Ingenix, okay, and the measure developer is here, and so, could you give us an introduction to the measure set?

DR. LYNN: Sure. Again, this is a measure that's been extracted from an application that tries to group all claims to episodes of disease.

Our approach with diabetes is to create year long episodes of diabetes, by gathering claims to the episode of diabetes, and then using comorbidities and what we call condition status factors for diagnostic information that's part of the diabetes episode itself, to do a -- create a severity score for that diabetes.

The measure then goes on, like all of our measures do, to use the severity of the diabetes to create expected values for our metrics that are all part of the measurement.

CO-CHAIR ROSENZWEIG: Okay, thank

you very much. Okay, yes?

DR. REEDER: I'm not familiar with Ingenix. Could you give me a time line? How long has this been going on? How rich are your data?

DR. LYNN: Ingenix is a -- has been around for 15 or 20 years. The product ETG is -- was originally a product of a company called Symmetry, which was purchased by Ingenix, maybe six years ago, and the product has been around for you know, 15 years.

DR. REEDER: Thank you.

DR. LYNN: That is the episode, the ETG product has been. You know, some of these other products that use it for these sorts of measurements are more recent.

CO-CHAIR ROSENZWEIG: And yet, Ingenix is a subsidiary of United Healthcare, is that correct?

DR. LYNN: United Healthcare is our sister. Our parent company is United

Healthcare, yes.

CO-CHAIR ROSENZWEIG: Yes, okay. Okay, so, this particular measure basically, to start out, is -- it focuses on the resources that deliver episodes of care with patients with diabetes, and they use a specific methodology that was developed by Ingenix, that's called the ETG methodology, episode treatment groups, and I'll get into how this is described, and I will probably ask our developer for more details, in relationship to this.

But largely, it's a grouping methodology that takes groups of visits and based upon an anchor visit, essentially creates an episode of care, and this is actually -- the applicability of this to diabetes is -- will be very interesting, because diabetes is such a chronic disease.

It actually is a mirror -- the review of this is, in a sense, a mirror of the previous Ingenix protocol that we started

reviewing, where there was clearly an event related problem, and this is quite the opposite.

This is a diabetes, which is a chronic disease, which involves lots of ongoing care, but they're able, through their methodology, to create episodes, distinct episodes of care for which they then are able to look at costs related to those episodes of care.

DR. HWONG: Jamie, one quick thing.

CO-CHAIR ROSENZWEIG: Yes.

DR. HWONG: And maybe also, with the measure developer here.

So, my -- in sort of reading this, in terms of the episode of care, because diabetes is, you know, classified as a chronic condition, it essentially is just one episode for the entire year, is that correct, or am I mis-interpreting that?

DR. LYNN: That is exactly

correct. We often use multiple years of data, of course, and then we create one long episode of diabetes, then go back and divide it into year long segments.

DR. HWONG: So, that sort of makes it conceptually a little bit easier to handle, right, that we don't --

CO-CHAIR ROSENZWEIG: Exactly.

DR. HWONG: -- have to worry about, you know, sort of these discreet episodes. It essentially becomes like a one-year, you know, service, you know, accounting of services.

CO-CHAIR ROSENZWEIG: Yes, in my review of this, it appeared to me that that one long year period was going to account for the majority of, certainly, the vast majority of the actual episodes.

But it didn't -- it wasn't clear that it would account for all of them. There seemed to be certain situations that might arise, where an episode of care could be

shorter than a year.

DR. LYNN: Right, so, what can happen, of course, the grouper itself can be configured in a number of ways.

But for the purposes of this project, it was configured in a way, so that all of the years end on the anniversary date of the end of the member's eligibility date.

You can -- depending on certain situations, you can configure it different ways, but that's how we configured it for the purposes of this project.

What that means is that the benefit is that your complete years tend to be at the end of your reporting period.

However, somebody could have, say, joined in June of one year, and then went through the end of the next, the following year. So, the last year would be a complete episode, but the year prior to that would be what we would call an incomplete episode because the member was only eligible for six

months.

And so, in this particular analysis, we have different folks that treat that different ways, and some people try to create -- try to basically, use that incomplete episode, but in this particular analysis, we did not do that. We only included complete year long episodes.

CO-CHAIR ROSENZWEIG: Okay, that clarifies it, actually. So, with respect to the first IM1, the summary of the evidence for high impact, this, I think, actually is summarized reasonably well, indicating that diabetes is an important disease, as I would certainly think so.

It involves a lot of patients, and they actually, also, do some analysis of their benchmark data from their own organization of about seven-million individuals, non and elderly, so that the diabetes represented 4.5 percent of the total population, of their group.

But that's largely because they were non-elderly. If you include the elderly, it will go up to like nine percent.

And they were able to look at, actually, the total cost per member per month, for these individuals was actually by most criteria, looking at other populations of people with diabetes was quite low, and that is largely probably because of the younger age of these patients, I would assume.

If you look at the average cost for patients with diabetes nationwide, it's much higher, either that, or just Ingenix is doing a good job, in keeping the costs down.

DR. LYNN: Let me make a comment about that, actually, because -- well, I just want to -- this is not just Ingenix data, by the way.

CO-CHAIR ROSENZWEIG: Okay.

DR. LYNN: This is not just United Health Group data.

CO-CHAIR ROSENZWEIG: Okay, so,

it's not just your own clients. It's a much larger database that you're looking at.

DR. LYNN: Right, so, Ingenix, it basically has a deal with all of the clients of ETG, that for decreased contracted rate, they share our -- their data with us.

So, it's all of our clients' data, not just United Health Group.

CO-CHAIR ROSENZWEIG: Okay, and they actually give some data, as well, on the number of prescriptions, costs per -- specialty visits, and various other encounters, as well.

So, I thought this was pretty well presented, okay.

The next section is opportunity for improvement. This particular section largely talks about the fact that there is fairly significant variability -- well, actually, no. No, that's the next section.

So, this basically is a fairly short section that indicates that obviously,

that there are lots of costs associated with diabetes, and that includes the ability to lower costs.

But actually, doesn't give much specific rationale for it, but indicates that this kind of methodology might be able to help with that.

Okay, and then the next section that describes the summary of the data, showing variation across providers.

Now, in this particular section, they are largely looking at variation by geographic areas, and indicating that certain areas have much more resources available, and that there seems to be a correlation, at least with respect to care that -- areas where there are more resources available tend to have more costs, more resource utilization, which has been demonstrated in a number of disease states.

I wasn't aware that this was -- I don't know if this is specific to diabetes or

not, at least in their discussion here, entirely. It's mostly chronically ill, patients who have chronic illness, in general.

DR. HWONG: Right, I think the only one, in terms of the references. So, I agree, it's like very broad.

I think there was one, in terms of the ambulatory care sensitive conditions, where one of the highlighted, at least in the blurb, one of the highlighted conditions is sort of looking at poorly controlled diabetes, and sort of the utilization rates of like hospitalization and ER afterwards.

So, I thought that was probably the only one that was very specific to diabetes, at least from, again, the quick review of the blurbs, I can't say, you know, if you dove deeper into some of these other ones, they break out diabetes or not.

CO-CHAIR ROSENZWEIG: Okay, so, but it largely discusses it in comparison to specific -- into geographic areas and into

areas where -- they are saying that certain geographic areas have high resources and others have low.

I would have liked to have seen a little more of a discussion of other issues, related to types of providers, issues related to variation -- on other issues than geographic.

But I thought it was reasonably well presented, as well, okay, and then they include citations for the variety of data, on their variation, and once again, using their ETG based condition, they come up with an observed to expected ratio of their costs per episode.

Do you want to comment on that, at all?

DR. LYNN: Well, I mean, eventually, the measure will look at other metrics besides costs and utilization metrics, as well, ER visits, hospital days, I think.

CO-CHAIR ROSENZWEIG: Okay, and

then there is a summary of the discussion of disparities by population group, and this particular area also discusses efforts to improve healthcare delivery in various areas.

It doesn't really specifically discuss underserved populations or socio-economic issues. It is mostly looking at areas where expenditures are higher versus other areas related to overuse, misuse and waste.

So, the focus, I thought, was very much related to that, rather than other issues.

MS. PARKER: Yes.

CO-CHAIR ROSENZWEIG: That I thought could have been included.

MS. PARKER: Right, I think in this section, they kind of fall short, and they acknowledge that there are disparities, but they don't really go into what those disparities look like.

It's kind of a more general

discussion on yes, they exist and they exist here, but they don't go into it.

So, I think again, it could be improved, to support the need for the measure.

CO-CHAIR ROSENZWEIG: But the focus a little more then, on the -- in the other ones, as it related to efforts to try to eliminate waste, duplication of services, things of that sort, which is perfectly reasonable, it's just a little different in its focus than some of the other proposals that we've had.

Then, the measure rationale for analyzing variation, basically, they say that they want to reduce unwarranted variation and eliminate unnecessary services, but they don't really -- and but they don't really describe how the measure relates to this, as much as sort of the use of robust -- as they say, a robust approach, including medical homes, value based payment and accountable care organizations.

So, they see this as a foundation for the use of those kinds of approaches.

So, it's not very specific, at least, with respect to the rationale for analyzing the variation, at least from my point of view.

DR. HWONG: Jamie, I had the impression, again, not so much about like diabetes, per se, but sort of the two things is, you know, allowing, you know -- having this sort of assessment, to allow sort of this, you know, classification of efficiency of, you know, providers and sort of that second blurb down there, I felt like it was interesting, here, is -- you know, Ingenix, in terms of this ETG grouper methodology, you know, it says that you can use the output on individual providers, roll that up.

And you know, so, this is one of the things that it's not like it's for a health plan. It's for provider group. It's for an individual physician. It's sort of

saying that it actually should be able to serve you well, in all of these categorizations, from individual, up to provider group, up to full delivery systems, you know, in particular, I think there is this focus on ACOs and you know, how that may be more relevant moving forward, right, in terms of having these types of statistics for those groups.

So, you know, I thought that was interesting, a little different --

CO-CHAIR ROSENZWEIG: Yes.

DR. HWONG: -- you know, the intent of the measure.

CO-CHAIR ROSENZWEIG: Yes, we get into that a little bit later -- go ahead.

CO-CHAIR CURTIS: Yes, I was going to say, but I think this is where I picked this up on the AMI measure, is that would measure developer one have to consider this from one or two of these perspectives, or do you want to get us to consider it across the

broader spectrum?

And I think it's relevant because the sensitivities at the physician level may be different than they are at the payer level or ACO level.

CO-CHAIR ROSENZWEIG: They're currently in use to evaluate providers, provider groups, and health plans, as well, all three. Is that correct?

DR. LYNN: That is correct, employers, although, at the level of providers, it's you know, aggregated with other diseases, not just diabetes.

CO-CHAIR ROSENZWEIG: Okay, and then the next section, the resource use categories, I did review the additional table, and it looked like that, in fact, it was a pretty comprehensive list of a whole variety of different categories that they included. It's a very robust huge list, in fact of various categories that they use for evaluation of the criteria, and it certainly

looked adequate to me.

I didn't know what all the numbers were, frankly. I mean, they don't -- you don't categorize them by ICD9 codes. You have your own map codes, for all these various services.

DR. LYNN: Yes, this is --

CO-CHAIR ROSENZWEIG: That are different from ones -- the ones that we normally use, but they're very, very extensive.

DR. LYNN: Yes, it's a roll-up of procedure codes.

CO-CHAIR ROSENZWEIG: Yes. Okay, so, do you want to stop here and vote on those, this whole group of measures?

So, with respect to the first one, the importance of the measure, summary of evidence of high impact, I basically -- that is 1a, I don't know if we have to read what that means, after doing this over and over again.

But I gave this a high score. I thought this was reasonably well presented.

(Off mic comment.)

CO-CHAIR ROSENZWEIG: You'll never forgive me for that. At least some people have selective memory.

Okay, all right, so, everyone -- okay, I'm glad that we have agreement on that one.

Now, with respect to the benefits envisioned by the use of the measure and the opportunity for improvement, I thought their case was a little bit skewed towards dealing with issues related to duplication of services, unnecessary services in regions and making too much medical care available.

And I guess because of that, and because it didn't consider all of those other issues, I gave it a moderate score. But I think it was adequate, let's put it that way.

Okay, three high and five moderate, okay.

Now, the demonstration -- the data -- the next one is --

MS. TURBYVILLE: One-C.

CO-CHAIR ROSENZWEIG: -- 1c, which is --

MS. TURBYVILLE: Measure objective.

CO-CHAIR ROSENZWEIG: Okay, yes, the measure -- the construct -- the objective is -- and the construct for resource use costs are clearly described.

I didn't think they were that clearly described in this particular. It was kind of a fairly brief description. So, the purpose, I guess, is described reasonably well.

So, I have given it a low reading, but the purpose is described in a fairly short manner, so, I probably -- I think I'd probably move that up from low, to moderate, frankly, from my recommendations, with respect to this particular section, after thinking about it a

little bit more.

Okay, so, three high and two -- and five moderate, okay, and then, I did believe that they were -- they had a really quite extensive and complete -- with respect to 1d, the objective and resource use and construct for resource use are clearly described.

I thought they actually did a good job, of summarizing that. I gave them high marks on that one.

Okay, good, all right, eight high. Now, we'll move onto the scientific acceptability, yes, scientific acceptability of the measure properties, the extent to which the measure produces consistent, reliable, credible valid results.

Basically, they described their methodology, with respect to the foundational of the episodes of care. It's a different methodology that we've encountered, with respect to the other protocols.

We actually had a meeting about a year or so ago, that I attended, that was sponsored by NQF on the whole -- how to define the diabetes episode of care. It was actually an interesting issue, because it's so difficult to be able to come to agree to a common, sort of -- how to interpret the episode of care because of the nature of how diabetes is cared for in the fact that care of patients with diabetes is so shared among multiple providers.

So, it seems that although they use the episodes of care methodology, they're largely really talking about a time based, from what you're telling us, a year long -- in a sense, even though the methodology is different, you're coming, basically, to the analysis of a year long grouping of diabetes related costs, much like the other protocols that we've encountered. Is that correct?

DR. LYNN: Yes, that is correct.

CO-CHAIR ROSENZWEIG: Yes, okay.

CO-CHAIR CURTIS: Just so, I think -- just had to clear my head.

So, for the vast, vast majority of patients, the episode is a year. There could be instances where it would be slightly less than a year or is it 100 percent at a year?

DR. LYNN: You could have episodes that are less than a year, but for the purposes of the measures that are at the end, those get eliminated.

CO-CHAIR CURTIS: Got it, thank you.

DR. PALESTRANT: Can I just ask a question about the general methodology?

MS. TURBYVILLE: Is that David?

DR. PALESTRANT: Yes, it is. David Palestrant.

MS. TURBYVILLE: Go ahead, please.

DR. PALESTRANT: Yes, the issue I allude to the other measures, but and they're all exactly the same, with respect to the verbiage, and let's say, very impressively

written. It could almost be a textbook, in terms of the different issues that come up.

The question I have is, this specific methodology, has it been validated externally, in the literature? It seems like it's proprietary, but my question is, I guess, has this been scrutinized outside of the company?

DR. LYNN: Not really sent the methodology outside of the company to be validated, per se.

The methodology is made available for folks in academia, to use for various studies, some of which are around the episode grouper, itself.

So, the short answer is, probably no, but it had -- it's obviously, used by a lot of entities external to Ingenix, some of which are academic.

DR. PALESTRANT: So, some of the experts in the panel in this area -- can you comment on this methodology, or do you have a

comment on the -- at the end, about what you think?

DR. HWONG: So, when I was looking a little bit at sort of trying to understand, because it's a fairly complex system, right, this ETG methodology?

CO-CHAIR ROSENZWEIG: Yes.

DR. HWONG: You know, just looking around, getting some background information, but I want to say, CMS, there is, you know, I have this article, but CMS, in 2008, you know, did an extensive study on ETG's versus MEG's, two proprietary systems, in terms of, you know, evaluating kind of like the differences.

And it turns out, I mean, you know, the article is totally not particularly relevant to this, but you know, in terms of like, there are sort of just subtle differences.

So, in the sense of, just to answer that question, I think, you know, this problem has been around for a long time, and

there have been, you know, public entities that have, you know, evaluated that for their purposes and compared it to other existing, you know, grouper of methodologies.

So, I think there is some level of familiarity, you know, with that in the external space.

CO-CHAIR ROSENZWEIG: You turned me off?

CO-CHAIR CURTIS: I think it's easier to hear the phone, when the microphones are off. I think if we adjust the volume and --

CO-CHAIR ROSENZWEIG: Okay, I see, I have a sensor next door to me, okay, all right.

Okay, all right, so, they basically take a fairly wide range of data, including, you know, basically, a claim on -- claims, they use diagnosis and NDC codes, HCPC's, ICD9, CPT -- I don't even know what NUBC revenue codes are, I'll have to be

honest.

DR. LYNN: Those are the hospital codes that -- the line items.

CO-CHAIR ROSENZWEIG: Okay, and even non-standard other local codes are taken in and are cross-walked, and actually, added to valid comparable codes.

And they look at a wide variety, including in-patient facility, out-patient facility, pharmacy benefits and a variety of other things.

They are fairly -- and they list a whole -- a number of them on page 12, okay.

They're fairly comprehensive, in terms of all of these features, but they're also fairly -- the data inclusion is fairly -- I mean, they're basically, fairly specific for diabetes related, in a more narrow sense than certainly, was given to us for the NCQA and even -- and also, it's more narrow than what was given to us for the ABMS.

They're more focused on clearly,

treatment of diabetes related -- treatment of diabetes itself, as opposed to all of its complications, am I correct in that?

DR. LYNN: Yes, that's correct. Again, you know, this is an extraction from a methodology that groups into many different diseases and conditions, and you know, our philosophy is always -- has, for the most part, has been, you know, you can put things together, but it's hard for folks using the product, to take them apart.

So, we do look at diabetes in a narrow way, and you know, we have folks that use it in a broader way, and then include multiple episodes related to diabetes to do that.

But then we have other folks that would say, "You know, well, I don't want to include diabetic retinopathy in there, because I want to be able to pull that out, and look at how my opthamologists are handling that separately."

DR. HWONG: You know, I got the sense in -- I'm sorry, I got the sense in my review, you know now, that we've seen sort of the three different measure developers, right, you know, NCQA, clearly, the broadest.

You find the diabetics and you throw all the -- the claims associated, or services associated with the diabetic patient.

The ABMS versions are -- they get down to be very specific, I felt like, in terms of, here are the meds, here are the, you know, E&M visits, that are associated, you know, with this diagnosis code, et cetera.

The Ingenix system, as far as I could tell, it felt like it was in between, for me, in the sense that, they have specific codes that have to be sort of the -- for the anchor or the primary diagnosis, but in terms of the actual episode of what claims get counted, in terms of the cost, you can actually start to associate a lot of things that, you know, wouldn't -- weren't

necessarily included on the ABMS, you know, criteria, in terms of being specific.

So, you know, probably this may come up a little bit later, if we looked at the data dictionary and some of the clinical logic, but there is some services in there, like, I don't know, like anesthesia, you know, and there is some kind of like, somewhat little bit random kind of stuff, that sometimes kind of gets captured in there, for better or for worse.

I mean, somewhere -- I'm saying it is somewhere in between. Clearly, we've looked at, you know, again, methodology that takes all claims, and then we look at things that are very sort of clinician picked and very focused on ABMS.

I kind of felt like this sort of fell somewhere in between.

CO-CHAIR ROSENZWEIG: Well, the range of the curves was broad, but it seemed like all of them had to be related to a

diabetes episode.

CO-CHAIR CURTIS: But that's my question, and that's where it becomes a black box, here, is that if you go through the spreadsheets, which are extensive, of whether or not a particular claim was -- the strength of association, I can't remember, I'm going to get the nomenclature wrong.

But that seemed potentially, completely arbitrary, and that is supported, in a sense, by the noise that you identified, Connie.

And so, that is the validation that I actually want, right, that's why I have a hard time evaluating the quality of this measure, without knowing and feeling confident that this was something that made clinical sense, and because where were so many episodes, where it couldn't make clinical sense to a group that -- with diabetes, or it just didn't make -- I couldn't figure out the clinical sensibility to it. That troubled me,

and that is true for all of the measures.

I think probably something we should take back to the Steering Committee, or I would propose, is that the Ingenix measures are, as you pointed out, very hard to disentangle and look at in isolation, and it's almost like you need an entirely different approach to look at Ingenix, and sort of the ETG grouper methodology, and evaluate it as a whole, as opposed to piece by piece, because I think we're going to run up against this same thing in all the measures.

DR. PALESTRANT: I just wanted to second what you said. I think that falls from these categories, which are chronic, so, one of the other ones I reviewed was coronary arteries, which I guess, the ongoing disease and could be episodic and also be chronic.

There are so many different areas where this would fit into and how to you make sure that this is, you know, that you're -- you need to be -- it was hard to get specific

on how that code is actually captured and defined, how that -- but what is the criteria, for actually measuring? What are the actual things that you're trying to measure?

CO-CHAIR ROSENZWEIG: Yes, I would agree with those previous two comments, and I was going to get to that a little bit later, with respect to the black box issue of this.

But with respect to the -- they have this section on missing data. There were some parts of this that I really just didn't understand.

When they said missing pharmacies data for some members and populations, pharmacy data can be missing, generally, due to different factors, including not having a pharmacy benefit with the entity collecting the data used for measurement or pharmacy services being managed by a pharmacy benefits manager for the measurement entity.

Where pharmacy data are not generally available for a member, adjustments

are required to ensure valid comparisons. What are these adjustments? I mean, I just didn't -- you know, and the next section, it said that in fact, the methodology didn't require pharmacy data at all, but somehow, there would be adjustments that would be thrown in there.

I just didn't understand, how that would work.

DR. LYNN: Yes, we're moving beyond the ETG, but it's definitely part of its measurement, which is, you know, how do you deal with a group of members, where some of them have pharmacy data and some of them don't?

And we've looked at a lot of approaches on this, but what we've come down to is basically, when you -- once you've grouped the data and you start to create the O to E ratios, the critical part in the denominator is, you know, what is the expected value and how is that sort of stratified?

And so, what we've done is, we've added to that stratification, whether the member had pharmacy benefits during that time or not, and then that drives the expected value, obviously, drives the expected value down, when they don't have pharmacy benefits and increases it when they do.

So, that is that approach that this measure has taken, to be able to combine folks that have pharmacy data with folks that do not.

CO-CHAIR ROSENZWEIG: But the devil, to a certain extent, is in the details, as to how -- you know, how that -- you know, how you compensate for that.

DR. LYNN: Well, I'd be happy to discuss the details.

You know, so, let's take a stratification of diabetes around how it might be done.

So, you might -- it's not -- it might not be how it's done in here.

You take a member, all of the cases around a peer group, that have diabetes, a peer group of providers or groups, or you can do it in a larger setting, and the -- the stratification would be, you go into that peer group and look at all of the episodes of diabetes, and they're basically eight stratifications, the four severity levels, and each one, whether or not they had a pharmacy benefit or not.

And therefore, you create the expected values, based on those eight buckets, and use those expected values, so, when you have a member -- when you have -- and I'm just using three, not that we would ever use three, but just sort of to be simple.

You had a member that -- a diabetes episode that was severity one with pharmacy benefit of the diabetes member -- episode, severity one without, and diabetes episode that was severity level two and had a pharmacy benefit, and then you calculate --

use the -- in the denominator, use the expected value from the appropriate strata, and that is how we account for that.

CO-CHAIR ROSENZWEIG: Okay, and then there is a fairly lengthy discussion in the protocol, under the clinical framework, which discusses how these ETG's are created, based upon anchor visits, anchor records, or -- and then episodes that are created from the anchor records, and then non-anchor records that are then grouped together to the episodes.

And then the co-morbidities and complicated factors are added, subsequent to that, or treatment of those issues are added, subsequent to that.

To a certain extent, this is moot because with respect to at least this diabetes protocol, you're really considering all episodes -- I mean, all events of care within a specific year. Am I correct, in assuming that?

DR. LYNN: Your assumption is right. I don't understand what part is moot, because of that.

CO-CHAIR CURTIS: Well, I think from my read, it was like, if you have an episode that's six months long, and comparing that to the resource use of a year long episode, would make comparisons difficult.

So, in that sense, it's easier or more intuitive to understand, since they are all at least one year, or they are all one year.

CO-CHAIR ROSENZWEIG: Yes, yes, that is what I was trying to get at, yes.

DR. LYNN: Yes, that is correct.

CO-CHAIR ROSENZWEIG: I mean, I still think, yes, you know, you may be able to stratify or be able to analyze the costs, based upon what goes to the anchor and what goes to the others separately, but to a certain extent, the total costs are all lumped together, as a part of this whole process.

Okay, and then there was about -- the issue of finalizing the episodes. What does that exactly mean? I just had a question about that.

It says, "Finalizing an episode of diabetes involves determining whether or not the episode is complete, assigning co-morbidities and conditions status factors and calculating a severity score and an associated severity level."

So, how are the severity scores and severity levels determined?

DR. LYNN: So, for each episode of diabetes, there are a number of markers that occur during the year long episode.

CO-CHAIR ROSENZWEIG: Yes.

DR. LYNN: There are co-morbidity markers. These are, in the case of ETG, these are episodes that occur outside of the diabetes, that have an indirect effect on the cost of the diabetes.

And then we have what we call

condition status factors, which are -- and these are all by the way, diagnostically driven, factors inside of diabetes that directly affect the cost of diabetes, because these are claims that are actually grouping to the episode.

And each one of those, we've taken these markers and we have, you know, put them in a -- a linear regression model, to look at the direct effects -- I'm sorry, the effect of the markers, as well as the effect of interactions of the markers, and as well as the demographic information.

And so, all the grouper really has to do -- once that difficult modeling is done, the grouper is just going to a table and saying, this marker adds this much severity to the episode, and adds up all of those scores, to create a severity score for the episode of diabetes itself.

Finally, we put episodes that have similar -- that have severity levels that are

similar into buckets, called severity levels.

So, we take that real number that goes from zero to five or six, or something like that and we divide it into buckets of four severity levels, one, two, three, four where low is the highest -- low is the -- low is definitely not the highest.

One is the lowest, and four is the highest, and then, that's how we create our sort of statistical unit right along with, whether you have pharmacy or not, to figure out what the expected value is inside that level.

CO-CHAIR ROSENZWEIG: Your basis for your severity scores are all based upon cost data, that you've accumulated?

DR. LYNN: Yes, that is a great question. You wouldn't be surprised, probably, to hear that if we have another insurance company that we get data from, that they don't give us the cost. They give us everything else, but not the cost.

So, we have a standard priced methodology -- this is for the actual modeling purposes, with a standard price process that goes through and standard prices all of the claims, and then uses the standard price cost as the dependent variable in the model, that is exactly right.

But because it is standard priced, the dependent variable is actually -- is more like resource utilization than cost, because the contracted rate has been taken out.

CO-CHAIR ROSENZWEIG: Okay.

DR. HWONG: Jamie, I wonder if we can go back, you know, in terms of the black box comment that you brought up earlier, and the same with Jeptha, in term of the -- again, I want to sort maybe get a little bit more clarity, in terms of the types of claims that get ultimately put into this one year long episode, right.

So, I think it makes a lot of sense, in terms of the primary, you know, the

primary anchor dates, no problem. You have a very highly specific list of codes for diabetes. You look at it and say, "That is diabetes. That is good."

The problem, or the concern, I mean, it's not necessarily a problem, but you know, is when you get to those incidental diagnoses codes, right, which can, essentially, they're ranked or not ranked, they're tagged as being specific, non-specific sign or symptom, right.

And so, if it's specific, right, to diabetes, it can get pulled in, it can get pulled into the overall evaluation of the cost of that episode.

So, this is the list. Again, primary diagnosis codes, no problem. That looks fine. This one has just some really funny things, in terms of the specific -- like, what you consider specific, and I'm just reading about that point. You can --

DR. LYNN: Yes, yes --

DR. HWONG: Yes, okay, maybe you can just --

DR. LYNN: Yes, because you were going down a little bit of wrong path.

DR. HWONG: Okay, good, yes.

DR. LYNN: So, the specificity is a description -- is a description of the diagnosis code, not the relationship between the diagnosis code and the episode.

DR. HWONG: Okay.

DR. LYNN: The primary and incidental is the relationship, and incidental can have a rank associated with it.

DR. HWONG: Okay.

DR. LYNN: But the specific is basically -- it's like -- it's describing the diagnosis codes.

So, you're basically saying, this diagnosis code seems to describe a specific disease, not necessarily diabetes, okay.

DR. HWONG: Okay.

DR. LYNN: And that a non-specific

is just trying to get a specific -- trying to get at describing disease, but it's non-specific, in other words, it could describe a number of diseases, and these are usually what they call mis-codes, right, the three and four digit codes.

DR. HWONG: Okay.

DR. LYNN: And then we have the signs and symptoms, that don't describe specific diseases.

So, then it's the relationship between that code and diabetes that's incidental, which means it doesn't have as much power to join the episode, as a primary diagnosis code.

But the diagnosis code itself has a higher priority because it's specific, as opposed to non-specific sign and symptom.

DR. HWONG: Okay, so, I guess, you know, and I think this conversation sort of highlights it, it is a little it's -- so, it's a little challenging to kind of wrap one's

brain, kind of around that, right.

I get the sort of, the primary diagnosis codes and I guess sort of, you're saying sort of specific/non-specific is different than primary versus incidental, like you know, they're sort of slightly different concepts there.

So, if you could explain, like, in terms of -- you know, I understand, you've got an anchor, right, you've got an anchor that comes in, primary diagnosis code, looks good, it's diabetes, whatever procedure gets counted in there.

What else gets put into that episode, then?

DR. LYNN: Then once that episode is started, then other primary diagnoses can join that episode, and there is a higher priority there.

Incidental diagnosis codes can join that episode, and the -- and again, this is where we're -- you know, we have a little

trouble, because we're sort of showing you how diabetes works, but it works in context with other diseases, in that these claims have to compete with other episodes that it could potentially start or join.

And so, we're sort of casting a wide net, from the incidental standpoint, and from the procedure standpoint, in order to try to drive -- because there is a competition going on, and you know, even if it's eligible to go to diabetes, it might not, and in some cases, probably will not, because there are other episodes that are competing.

DR. HWONG: Got you, so, when I look at this incidental diagnosis code list, this concept here is that, you know, it will be viewed in the full context of how many other primary diagnoses or ETG's or episodes are being -- you know, ETG's, separate ETG's are being opened, and you know, that code may go to the diabetes episode, or it may not, right.

So, and that's fine. That's just the system. It's just, you know, we're evaluating it, when I'm looking at it, when I saw it on the spreadsheet, you know, it was like, oh, is that specific/non-specific to this group?

You're just saying that this is this general bucket of, you know, services that generate cost, that in the end -- and so, this is where it was the little black box, but in the end, where it's kind of all weighed out, some will go to the diabetes episode, others will not.

DR. LYNN: That is correct, and that is why there is, you know, an extensive discussion of the tie breaking logic in this document, and I -- believe me, I know, I understand how that, you know, is difficult to sort of wade through.

But we tried to put it in there and --

DR. HWONG: It's not a bad thing, just it is a challenge, that's all.

CO-CHAIR ROSENZWEIG: So, with respect to your risk -- your severity of -- your severity score, your severity scoring system that you've developed, is it -- is the methodology for that available, to others, or is it proprietary for Ingenix?

DR. LYNN: Sorry, I don't mean to turn my mic on at the same time.

You know, it's proprietary. We've obviously shared it here, in great detail, greater detail than we -- I really don't know, to the extent that we sort of share it.

But it is proprietary. It was developed by us. You know, we haven't always shared the actual weights on the different markers.

We've always shared the markers, but we've only shared the weights, in certain circumstances.

CO-CHAIR ROSENZWEIG: Okay.

DR. MARWICK: Can I ask a specific question?

CO-CHAIR ROSENZWEIG: Sure.

DR. MARWICK: I'm still not sure I have my head around this.

Say I have a patient who is admitted to the hospital with diabetes and heart failure. That patient might end up going to the heart failure ETG, presumably.

If I have a patient who has a background history of heart failure, but presents with a diabetic problem, so that their primary problem is diabetes, I take from what you are saying that they will probably end up in the diabetes bucket.

But then they may not be terribly different entities.

DR. LYNN: That is true. You know, there is no question that when you have -- you know, the hospital admission is a little bit different because the hospital admission, when you look at the diagnosis code list, there is a meaning to the fact that some -- that the diagnosis is primary, the first

one. It's less clear cut, in other sorts of claims.

So, you know, the grouper, in the case -- in the special case of an in-patient stay, the primary ICD9 code drives, always drives where that episode groups, which is not the case in others.

They are equal, except for the order on the claim, as the final tie breaker.

So, you know, in the case of an in-patient claim, you know, it's, you know, really very consistently going to go to what was the primary reason for the admission.

I mean, I think it's a shared sort of problem that a lot of these things have, that someone who presents to the hospital with a primary diagnosis is diabetes, and the secondary diagnosis is CHF, might not be that different from someone who presents with a primary of CHF and a secondary of diabetes.

But you know, it's an issue that a lot of folks sort of struggle with, and we --

our current methodology does not really split up claims. We don't split claim costs in multiple episodes, although, we're actually -- not that this really matters, but we've always been worried about that, thinking about it, and looking at ways that you could divide up the costs.

In addition to that, in the cases where -- so, you know, looking like, how the claim lines group, looking at what the co-morbidities are, from say, an MS-DRG standpoint, and dividing up the costs that way.

Looking at the professional claims that occur during the hospitalization and seeing if that can help you divide up costs, when you do have these cases, where it's diabetes and congestive heart failure.

But the product right now, that you're evaluating, will take that cost for that admission and group it to a single episode.

CO-CHAIR CURTIS: I would say, it's precisely defined, as to what bucket it ends up in, in the black box is the appropriateness of that decision, right?

CO-CHAIR ROSENZWEIG: Okay, so, moving onto page 21, they talk about how the major condition factors that are defined for their diabetes, at least for their anchor, they have five categories.

You know, basically, they're talking about this very specific diabetes diagnosis, either diabetes Type 1 or Type 2, diabetic coma, which presumably, or hyperosmolar state or ketoacidosis.

I was wondering why you didn't include hypoglycemia, as -- which is certainly a -- would be an appropriate very specific -- you know, specific diabetes related complication, not a complication, but an element related to diabetes and I just was curious, if that should be included, as well.

Also, the co-morbidity factors are

very broad. So, basically, any complication of diabetes, which normally would be considered very closely related to diabetes, is considered as a co-morbidity, at least in this methodology.

DR. LYNN: Yes, we didn't look at hypoglycemia as a marker. I mean, I think maybe we should have, but we did not.

We do look at those other markers and this is -- as far as this, you know, the co-morbidities, they are broad, that we put a lot of things in the model. They all had -- there was obviously, a lot of things that have an effect on the diabetes.

But the co-morbidities, again, have an indirect effect, right, because they're -- the cost for that co-morbidity is captured in a different episode.

So, the effect it has on the cost of the diabetes itself, is not the obvious fact that it increases the cost to the patient. But the indirect effect that, you

know, because there is this other disease occurring, it's making the diabetes harder to treat, from a utilization standpoint.

CO-CHAIR ROSENZWEIG: Okay.

CO-CHAIR CURTIS: Let me just, I think it's related, correct me if -- I apologize if I'm wrong.

But the risks, or the severity levels in the identification of the co-morbidities is taking place concurrently with the -- within the episode, within the year, correct?

And the problem, or the question I have for you is, there is specific guidance from NQF criteria in 2b5, to be very specific, talking about how we're -- sorry, 2b4, that for risk -- and this gets into risk adjustment, that it's -- you're suppose to adjust for factors that are present at the start of care, and could not represent complications, and so, maybe I'm getting ahead of myself, here.

But I just want to ask for some clarification, as to why you took that approach, and again, it's very different than what we do for outcomes measures and to me, could -- is potentially, could be problematic.

DR. LYNN: You know, we do consider co-morbidities that occur during the measurement year, that don't -- they're not limited to those that occur before the measurement year.

DR. WEINTRAUB: Then per Dr. Curtis' point, how do you set the product -- to that point, how do you separate out complications from co-morbidity?

DR. LYNN: I guess we don't.

CO-CHAIR ROSENZWEIG: Yes, it looks like, I mean, like they say, example for co-morbidity groups for diabetes included ischemic heart disease, congestive heart failure, and COPD. Well, ischemic heart disease certainly is a complication -- can be considered more closely related to diabetes as

diabetic retinopathy or diabetic nephropathy or something like that.

COPD is something sort of often a different realm.

So, and certainly, there are other examples, like you mentioned, multiple sclerosis, so, even there, that would be a co-morbidity, but it's not sort of part of the diabetes care episode.

And I didn't quite understand, are you taking -- are you considering all of the costs for all of these things, or is it -- no, you're not?

DR. LYNN: No, it's a marker, right, it's a marker that has an indirect effect on the cost -- all of these markers have indirect effects on the cost of caring for the very specific diabetes episode.

CO-CHAIR ROSENZWEIG: Okay, all right, and you use Moody's examples, and actually, there is a lot of data to support issues related to --

DR. LYNN: Yes, and the --

CO-CHAIR ROSENZWEIG: -- co-morbidity of depression, being associated with increased cost, okay.

All right, now, I had already asked -- I guess, I didn't quite understand how the severity scores were elucidated, but according to what you are saying, it is basically weighted, based upon comparable codes in your database, that are associated with similar levels of cost, or were there other issues -- other elements that go into them, in addition to cost?

DR. LYNN: Again, the weights come from a model that uses standard price as the dependent variable.

CO-CHAIR ROSENZWEIG: Okay, all right, so, I'm going to move onto page 27, and you're listing a lot of resource use categories, and most of these seem pretty straight forward.

I wondered why you didn't include

diabetes education, at all, as one of the resource use categories that might enter into the picture here.

(Off mic comments)

CO-CHAIR ROSENZWEIG: We're on page 27 S9.7.

MS. PARKER: Jamie, that was actually a question that I had on the previous one that I missed in my notes, is that exact thing.

I wasn't sure if how it's coded or if it was even -- I guess it wouldn't matter in NCQA, since they kind of group everything together.

But I think that's a very valid point, that is something is billed for, it does happen, and it is part of the standard of care for patients with diabetes.

CO-CHAIR ROSENZWEIG: It's definitely part of the standard of care, and it is billed for, the actual amount of diabetes education is -- that is billed for

between plans is very variable.

So, but it actually represents, you know, a cost component that should be taken into consideration, with respect to episodes of care, and I suppose, you know, whether or not it -- you can subsume it under evaluation and management services, I'm not sure.

DR. LYNN: We can definitely pull that out, as a separate category.

CO-CHAIR CURTIS: But you are suggesting that it's already in there, it's just not broken up?

DR. LYNN: I think that is correct.

CO-CHAIR CURTIS: Okay, so, if you could just check back with us.

DR. LYNN: Yes.

CO-CHAIR ROSENZWEIG: Okay, all right, and they describe in quite detail, the various -- how they define the various types of services.

I'm not sure I need to get into in depth, in describing it to the committee, but it also goes into -- I mean, once again, it -- I don't pretend to fully understand how these calculations are done, in order to create the scores, okay.

CO-CHAIR CURTIS: But the top line message here is that only the costs that are associated with claims that have an adequate -- - that are mapped to diabetes get captured, correct?

DR. LYNN: Those are the only ones that get captured in the episode.

CO-CHAIR CURTIS: In this particular --

DR. LYNN: Yes.

CO-CHAIR CURTIS: Right.

CO-CHAIR ROSENZWEIG: That is why I -- CO-CHAIR CURTIS: So, the admission for heart failure, that's grouped to heart failure, would be invisible in this particular measure.

CO-CHAIR ROSENZWEIG: That is why I thought it was --

DR. LYNN: That is correct.

CO-CHAIR ROSENZWEIG: Yes, that is why I thought -- yes, that was the basis of my interpretation, that this was more diabetes -- that the costs that were being evaluated in this were more diabetes specific and less related to total medical costs, than in other situations, that you were trying to actually eliminate costs for the variety of co-morbidities, or a lot of co-morbidities.

However, some of the co-morbidities do affect your diabetes costs, statistically.

DR. LYNN: Right, so, the ones that are outside of diabetes are markers for the severity of the diabetes, itself.

But that's correct, we're looking at the direct cost of diabetes, again, you know, if we have folks that want to analyze that unit, they can. If we have folks that

want to analyze the aggregation of diabetes and all of its sequela, they can add our episodes together, to do that.

CO-CHAIR CURTIS: So, one clarifying question, then. Does the intensity of coding variations, by region, physician, whatever, influence this, in terms of that tie breaking methodology?

So, I would assume that the more codes you have, or that, you know, how many ICD9 codes I check off, provides a different set of potential number of episodes that it could be attributed to.

So, one might be heart failure. One might be diabetes. One might be CAD, and so, I could see that there would be problems, based on that known variation and just the number of codes that are submitted.

DR. LYNN: So, two comments about that. We had done studies about looking at, you know, claims that have three ICD9 codes on it versus four, versus two, versus one, and

when you go from three to four, you're only changing grouping, like, less than a percent of the time.

So, it doesn't have an effect on grouping. But it could have an effect on co-morbidity identifications and markers.

But it only takes one diagnosis code to mark a co-morbidity. So, you know, that effect is relatively small, too.

CO-CHAIR CURTIS: So, for future applications, it would be very good to have that information, because persistently, in all my evaluations of the Ingenix measures, that was the biggest concern I had, is that stability of the assignment.

DR. LYNN: The stability of the assignment of the claim to the episode?

CO-CHAIR CURTIS: Right, that again, if I happen to click on heart failure one day, and heart failure and diabetes, the next time I see the patient, because I have two more seconds to think about what I saw the

patient for, you know, just that possibility of arbitrary assignment, or maybe I'm perseverating, so, I'll stop.

DR. LYNN: No, I think, you know, I -- we can probably -- we actually probably have that data someplace, because I know we've done that before.

I don't know if this is the data you're looking for, but you know, how do things change from one diagnosis to two to three to four?

(Off mic comments)

DR. PALESTRANT: Can I ask just ask one question, please?

DR. WEINTRAUB: He's just trying to show who is boss.

DR. PALESTRANT: Sorry, can I just interrupt for one second?

MS. TURBYVILLE: Please.

DR. PALESTRANT: Yes, so, say the one of the issues I don't quite understand is where -- it was in this -- this example, there

is the specialty care service, and there is the Excel spreadsheet that goes through all the different potential specialty services that could be attached.

Some of these things, I would not think that it -- it's basically, every specialty service that could be offered to any patient or any time, not diabetic specific at all. Some of them would be, of course, but some of them wouldn't.

So, I mean, to the trauma codes, which I don't think would be probably be due to diabetes, unless someone became hypoglycemic and drove their car off the road.

So, I'm not quite sure if we get into the value -- because I mean, what concerns me here is, there is suppose to be value -- resource use per episode, but it seems to be almost resource use per patient, because there are so many different attached episodes, so many attached episodes to each diagnosis.

DR. LYNN: What spreadsheet are you looking at?

DR. PALESTRANT: I'm looking at the Excel spreadsheet, it's the one that's in the red posting, that's 1595, and it's line seven, which is what I think is being referenced on page 30.

DR. LYNN: Yes, this is a comprehensive list for -- that's used for all measures. So, it is þ- includes a lot of stuff that's not related to diabetes.

CO-CHAIR CURTIS: But what I did see for the AMI measure, previously, was that it had the more specific assignments -- that's a bad term, but it had the assignments and I think it was S5, which I didn't see in this web-based.

So, there may be a missing spreadsheet, that might be more specific to this measure.

DR. PALESTRANT: The specialty code services, if I'm reading it correctly, is

-- because this is actually in each of the other measures for Ingenix, and it is this broad, unless I'm not understanding it, that these are all included, which makes me think that it's services per year, per patient, not services per diagnosis.

DR. LYNN: It's broad because it's used across all of these episodes, not just diabetes.

When it would be used for diabetes, only those procedures that were related to these categories would be included.

CO-CHAIR ROSENZWEIG: The way I interpreted it was that that big table, and in deed, the fairly lengthy listing of specialty care services and radiology and so forth, that only -- you know, that -- this is their entire list of things that they actually collect data on, but only certain percentages are actually -- only certain ones of these are going to be actually specifically assigned to diabetes.

You're not going to have too many

allergy tests or -- that are going to be assigned to diabetes, unless they have insulin allergies, for example.

DR. PALESTRANT: Yes, that would be the assumption, but I'm not sure we can make that assumption.

CO-CHAIR ROSENZWEIG: Okay.

DR. LYNN: We can narrow this list down specifically for diabetes, if you'd like.

CO-CHAIR ROSENZWEIG: Okay, it's just -- yes, it's a huge list it seems fairly generic, not just specific to this particular protocol.

All right, and then getting onto page 32, is that -- I'm just moving ahead a little bit, here.

There is a fairly lengthy discussion -- well, it's a discussion of their -- of how they use the risk adjustment method, to compare -- basically, they have their severity of illness system, which as they indicate, is proprietary, but then they use it

to be able to compare different providers.

So, whereas the previous material was discussing mostly comparing large groups, this is -- they can use the risk adjustment methodology to compare providers, as well.

They have, you know, Dr. Jones and Dr. Smith, and Dr. Jones is more expensive than Dr. Smith.

DR. WEINTRAUB: You know, it's always going to depend on sample size and probably for diabetes, it probably can be done, given the relative þ- frequency of diabetes and the distribution of costs are quite clear.

Do you know what the R-square is of your model?

DR. LYNN: No, I've heard you ask the other -- I've already written that in my notes, to bring the R-square for this severity model.

DR. WEINTRAUB: Okay.

DR. LYNN: Yes, I don't know what

you're looking at with the physicians, comparing.

CO-CHAIR ROSENZWEIG: Okay, and then on S10.2 stratification method, it says ETG stratifies episodes by intensity of service or total cost. Is it both or one or the other?

DR. LYNN: So, if you use -- if you re-price the data set that you're doing the study in, or I should say standard price, my boss gets really mad at me, when I say re-price, standard price, the data set, then what you're looking at is intensity of service, because you've taken out the contracted rate, and if you actually use the actual cost, then you're -- then there is total cost, which includes not only the utilization, but potentially higher contracted rates.

So, you know, depending on what you're trying to do, you would use either one of those methods.

CO-CHAIR ROSENZWEIG: Okay.

DR. WEINTRAUB: I don't understand what you're trying to do here with this stratification.

You said a bottom line is -- the severity level can then be used to stratify episodes by severity, measured as resource consumption. I don't understand what you are doing.

DR. LYNN: Yes, so, again, this is where we're assigning the severity level to the severity score.

So, the severity score, which is a real number, maps to a severity level, where -- that's the one through four, thing.

DR. WEINTRAUB: So, are you developing -- are you developing models by severity? Are you really stratifying, or are do you mean something else than stratifying? I suspect you don't mean stratify.

CO-CHAIR CURTIS: This is risk adjustment methodology.

DR. LYNN: Yes.

CO-CHAIR CURTIS: Like, there is no other -- essentially, right?

DR. LYNN: There is no other stratification, yes. So, that -- okay, what we're using to stratify is the severity score.

So, there is no clinical stratification.

DR. WEINTRAUB: So, you're actually not stratifying, using -- you're looking at -- what you're saying is, we can look at severity.

You can look at it like a sub-group -- what it is, is you're developing a separate model.

DR. LYNN: No.

DR. WEINTRAUB: So, you're not truly stratifying. They're really sub-groups.

DR. LYNN: Okay, they're not clinically stratified.

DR. WEINTRAUB: No, I didn't mean clinically. I mean, by statistically. I mean, in terms of modeling, they're not really

strata.

DR. LYNN: Let me tell you what they are, and you can tell me whether they're really strata or not, because I --

Again, you take a severity score of an episode and then each of these severity levels is a severity score that maps to a range, and you know, from zero to .5 is severity level one, that is what is being done, and if that's not strata, then it doesn't belong here.

DR. WEINTRAUB: So, what you're really doing is, what you're saying is, we can look at different sub-groups, by how severe they are, and look at -- and develop O to E ratios for those separate sub-groups.

DR. LYNN: Right, but you -- but also, you can combine O to E ratios, right?

DR. WEINTRAUB: Sure.

DR. LYNN: Okay.

CO-CHAIR ROSENZWEIG: All right, so, going on to S11.1, the attribution

process, it sounds like you basically look at all the various visits and you assign a specific provider to each of them.

Now, you're talking about physicians here, only. Diabetes is specifically a condition which a large portion of visits are actually done by people other than physicians.

So, I assume you're talking about all sorts of providers, and not just physicians, is that correct? Like, nurse practitioners, PA's, diabetes educators, podiatrists, things like that.

DR. LYNN: Yes, so, for the purpose of this, with this project, there are, you know, inside of the grouper, you can actually map different specialities to different -- to whether they're sort of considered ancillary or clinicians, and for -- when we did this, nurse practitioners and PA's were included in the clinician grouping, which would have the power to create an anchor.

But the nurse educators and the diabetes educators would not have been, and part of that may be a limitation to identifying that specialty in the data that we had.

CO-CHAIR ROSENZWEIG: I see, so, that could create problems, to a certain extent, because the way you're setting this up is, at least as I understand it, you know, with respect to attribution, is that obviously, it's hard to -- you know, as I said, personal diabetes may be seen in the course of a year, by eight or 10 different providers.

Some of them may be NP's, working with the primary care doctor. Some of them may be NP's working with a specialist.

But I assume, the way you're lumping all of the primary care guys together, as a group, okay, so, suppose a person sees more than one primary care doctor in the course of a year, they would be lumped

together as a group and then, at least as I interpret it from this fairly lengthy discussion here, and -- that you would lump all of the, let's say, endocrinologists together, all of the cardiologists together, and somehow, attach the NP's to each of these providers, or would the NP's also be sort of a separate group, as well?

DR. LYNN: The nurse practitioners would be a separate group, although we don't -- they're not commonly evaluated, but they would be a separate group, and again, you can -- you know, inside the grouper, you can not group the nurse practitioner to the physician assistants, to clinicians, and you get a slightly different result.

But we want to create, you know -- they should have the ability to create the anchor. So, we're not trying to aggregate the PA's to the primary care provider. You can, of course, do this at a group level, in which they would be aggregated to the group level.

CO-CHAIR CURTIS: So, it seems like you've created rules for attributing to individuals as one option, but you're retained the option of rolling it up to groups or other payer, or other levels, right, and then actually, I thought the one that was most appealing was actually the panel approach, where you sort of assign, you know, within that type of payer system, if you have a PCP assigned, you could attribute everybody and that information is, in my opinion, most actionable.

DR. HWONG: Yes, I mean, what I liked about, in terms of this measure, compared to let's say, the ABMS, and it's fine, I mean, ABMS has sort of one attribution logic, right, and you know, has its positives and negatives.

But this, it looks like you have, you know, four different kinds, right. You can sort of specify, you know, if you want it to be a PCP attribution, like who was just

identified as like the gate keeper in some ways, or who is the MD who has the most cost, who has the highest number of clusters, you know, within an episode, and or who has just the most face-to-face visits.

So, you know, I don't think they're not that -- it's nice to have that flexibility, right, to be able to sort decide kind of how -- what you feel like is -- you know, who you want to call responsible, let's say, for that episode and for those costs.

So, the one question I have, though, so, even with all those four variance, right, the episode only ever gets attributed to one -- like, a given episode and a cost for that, only ever gets attributed to one provider. I mean, you can roll them up, but like it's responsible by one provider, right?

DR. LYNN: That is correct, we don't -- the methodology does not divide the responsibility of episodes across multiple providers.

DR. HWONG: Okay.

CO-CHAIR ROSENZWEIG: That's within a peer group.

DR. LYNN: There is a possibility that rarely, if the peer groups have two different methodologies for determining the responsible provider, that very rarely, the episode would occur in one peer group and another.

But those groups would never be compared to each other. So, you wouldn't really be double counting the dollars, and it's very rare, and again, only if you assign different rules to different peer groups, as far as attribution goes.

If you don't assign different rules to different peer groups, then it won't happen.

CO-CHAIR ROSENZWEIG: Okay, so, it's more than just those four categories, you're dealing with -- within those categories of physicians, you know, on the physicians, on

an individual level, you're dealing with physicians in different peer groups.

So, it gets very, very detailed and very granular. You're dealing with cardiologists, primary care doctors, at least as I interpreted S11.2, you know, cardiologists, general surgery, and so forth and so on.

DR. LYNN: Yes, let's talk about that. I don't know if this is -- when we create the peer groups, part of the exercise is, you know, mapping episodes that are related to that peer group.

So, you know, if in the broader context, you know, if you fall down and break your foot, and your next door neighbor happens to be an endocrinologist, and he ends up with a foot fracture episode, we don't assign that to him, even though he may have been the responsible provider.

So, there is a map inside here that says, this peer group is responsible for

these episodes, and even if they get an errant episode outside of their area of speciality, it's not used to evaluate the provider.

So, general surgeon would not be in that map for diabetes, and cardiologist, you can debate it, but you know, I think usually, it's not.

CO-CHAIR ROSENZWEIG: Okay, but you say internal medicine, cardiology or general surgeon within a certain geographic area are examples of a peer group.

CO-CHAIR CURTIS: I think that might apply across the measures, as opposed to being specific for this one.

DR. LYNN: That is correct, and it should have been more specific, but it's something we missed, when we created specific documents out of that.

CO-CHAIR ROSENZWEIG: All right, okay.

DR. LYNN: So, I apologize for that.

CO-CHAIR ROSENZWEIG: That's okay. All right, and then sample size, or outliers, you do the Winsorization, like everyone else does, and then, with respect to sample size requirements, could you explain what you -- why a sample size of 30 is chosen?

DR. LYNN: Well, I think, you know, the sample size of 30 is --

MS. TURBYVILLE: Microphone.

DR. LYNN: Thank you. The sample size of 30 is really, you know, it's what is used, but it's not the important part, as far as we're concerned.

What's important is that you show a statistically significant difference to the threshold.

We used 30 because, you know, numbers lower than that start to get sort of ridiculous, even if you're statistically significantly different, and you have five cases or 10 cases, you know, I'm not sure, how really meaningful that is.

But the important thing is that, you know, whether you have 30 or 100 or 50, that if your score is statistically different than the threshold, then that's what should matter.

DR. WEINTRAUB: Well, you really should assign that in advance, rather than saying, well, now, it's statistically significant, but the problem then is that if you find something that's a trend, what do you -- how do you handle that?

So, the right way is to think in terms of power, and are you going to have enough power, with 30, to see a difference, if there is one?

DR. LYNN: Yes, I think, I mean, that's a good point. I mean, I think you could go back and say, "We're looking for differences that are -- you know, we want to find differences that are 25 percent or 20 percent, or something like that," and therefore, you would need a certain number to

do that.

DR. WEINTRAUB: That's correct.

DR. LYNN: Although, you know, there are -- the -- the issue that I guess, we have when we look at this is, suppose you're looking for differences that are 25 percent, and you -- and therefore you want to pick -- you would pick a number, like 50, but they have some providers or provider groups, that maybe have 35 or 40, but are sort of way out there, and you know, are statistically different from your threshold. Wouldn't you want to include those?

DR. WEINTRAUB: One way of doing that would be not just to have one number for power, but we could find a 50 percent difference at 30, and a 25 percent difference at 100, or whatever.

But I have trouble just saying, "Well, if it's statistically significant, it's statistically significant," and what do you do with the next guy, when there is a trend, when

you haven't set up your rules in advance?

CO-CHAIR CURTIS: Right, and so, I think we're getting a little off target, but I think it's an important point, in terms of the public reporting and the interpretation and use.

But I think we're probably unresolvable at this situation, but I would ask you to sort or take that under advisement, to sort of define what is clinically significant, before you -- in terms of the differences in cost, and that's something we haven't ask any other measure developer to do.

So, it would be sort of a little unfair.

CO-CHAIR ROSENZWEIG: But most of them haven't come up with a specific number, like 30.

CO-CHAIR CURTIS: Well, NCQA had the 400.

CO-CHAIR ROSENZWEIG: Four-hundred?

CO-CHAIR CURTIS: I mean, it really has to do -- it has to do more with the -- now, I'm getting philosophical again.

It's the statistical property of reliability, which is different than the other types of reliability that we've talked about, and that is testable, to say, "Okay, well, this is the number that you have to have," again, sort of a more stable case mix, if you sort of randomly sampled from the universe of diabetic patients.

We don't know what that number is, but that would, I think, influence the appropriate number for making categorizations and comparisons.

CO-CHAIR ROSENZWEIG: Okay, and then on page 38 and 39, they discuss the bench marking process, which we've actually discussed already and is outlined in S10.1, in more detail.

Okay, so, I think we've gone through this section, up to 'testing and

analysis', so, maybe we should -- or let's see, should we continue from there, or keep going?

CO-CHAIR CURTIS: We should, yes, I think following our lead, go through 2a1 an 2b1.

CO-CHAIR ROSENZWEIG: All right, okay.

CO-CHAIR CURTIS: So, I think -- so, with regards to --

CO-CHAIR ROSENZWEIG: So, 2a1 is the measure is well defined and precisely specified, so that it can be implemented consistently within and across organizations and allow for comparability. EHR measure specifications are based on quality data set.

Well, it's probably defined within Ingenix. The question is, is it defined for us?

DR. WEINTRAUB: Hard to tell.

CO-CHAIR ROSENZWEIG: That's hard to tell, as far as I'm concerned.

DR. MARWICK: Is the fact that it's

proprietary a problem, in that respect?

CO-CHAIR ROSENZWEIG: I don't know what the rules are for NQF with respect to --

DR. HWONG: That's a really good question, right.

DR. LYNN: Let me clarify that. Folks that use this product do have access to the weights. The weights are actually on a website, and all of how the grouper works and all of that stuff is available. I was just wrong when I said that.

CO-CHAIR CURTIS: Well, let me -- but so, is it available to any clinician off the street, who is being measured by this methodology?

DR. LYNN: Yes, there is a transparency website for ETG, where you can get all this information.

DR. PALESTRANT: Is your database used for -- in other words, is this database -- is this system being used for measuring value for conditions, or is it being used for

other purposes?

CO-CHAIR CURTIS: You were kind of breaking up there. If you could repeat it.

DR. PALESTRANT: Yes, it seems to me that this -- the database and this system is designed to measure the general usage per patient, and maybe by physician, but not usage per diagnosis, which is what we're getting at.

So, it seems that, and I may be wrong, I need this clarified then, is this database being used currently, and your system, for what we want to do with it, in other words, use it in this case for measuring diabetes over the course of a year, currently?

DR. LYNN: We're trying to measure diabetes. We're not -- you know, the unit of analysis is -- even when you take on the entire grouper, episode treatment groups, the unit of analysis is not the patients. The unit of analysis is the episode of disease.

CO-CHAIR ROSENZWEIG: Isn't that correct?

DR. LYNN: Yes, right.

DR. PALESTRANT: And you're currently being used -- I mean, are you currently using it -- are you currently using a diabetes episode treatment group as a commercial product, and giving this data out to your subscribers?

CO-CHAIR CURTIS: So, basically, is it currently in use?

DR. LYNN: The product is in use, that people -- you know, occasionally use it to measure solely diabetes, although they don't usually do that in the context of measurement.

They use it when they use -- look at only diabetes, they're looking at, you know, employee costs or health system costs or things like that.

But we do -- you know, it is -- and the grouper, as a whole, is used in its different conditions, to do exactly this, and has been used for a while.

CO-CHAIR ROSENZWEIG: The question is, is it well defined and precisely specified, so that it can be implemented?

My concern would be the fact that I don't feel it's precisely specified within the structure of this.

I mean, it's specified in a variety of ways, but --

DR. HWONG: But -- oh, go ahead, yes.

CO-CHAIR ROSENZWEIG: Yes, go ahead.

DR. HWONG: You know, it's -- there is where I sort of have that, you know, a hard time with this, right.

But I think I'm leaning towards one direction. I think it's precisely specified, like, the product itself and understanding how it's suppose to work, and how you've laid out the logic and even -- you know, to the extent of transparency that you provided to us, you know, in the measures, I

get the sense, if I bought the product, and I was using it, I know exactly how that works and you know, I can kind of look these things up.

I think, getting to Jamie's point, where it might be tough is if we -- if someone didn't buy a product and I, with my development team, wanted to try and build this, right, we'd probably get like a good distance, but I think there would be some -- you know, like, so, in terms of, is it spec to this point, where I could sort of reproduce, you know, this? I think there is probably a little, you know -- it would take a little work, and you know, you'd have to sort of build that.

So, I think I'm sort of leaning towards -- you know, this is just me, just sort of, just for conversation, but like, in terms of, do I think these are precisely specified, you know, in terms of the use of the product, and this sort of system? I do

think that.

I think, sort of, this whole sense of like, you know, is it enough that some outside, you know, group or entity who want to try and build this, right, you know, could do it or reproduce this, right, you know, may be a little bit more tricky.

DR. WEINTRAUB: I think that's very well said. I think you got right to the heart of it.

CO-CHAIR ROSENZWEIG: All right, so, I don't know, I would give it a moderate score.

DR. WEINTRAUB: But if you bought into what Connie said, you would give it a high.

CO-CHAIR CURTIS: So, let's go ahead and vote.

DR. WEINTRAUB: Electronics?

MS. TURBYVILLE: We're missing one, now.

(Off mic comments)

DR. LYNN: I didn't vote.

CO-CHAIR CURTIS: Did Brenda?

MS. TURBYVILLE: Brenda Marie is not at the table right now.

CO-CHAIR ROSENZWEIG: Okay, the next one is 2b2, is that correct?

CO-CHAIR CURTIS: 2b1.

CO-CHAIR ROSENZWEIG: 2b1, okay, measure specifications are consistent with the evidence presented to support the focus of measurement under criterion 1b. The measure is specified to capture the most inclusive target population, indicated by the evidence, and exclusions are supported by the evidence.

To me, it seems like that in fact, yes, that the measure specifications are consistent with the evidence presented, and it certainly does -- it captures an inclusive population. It has a lot of data in it.

CO-CHAIR CURTIS: So, just my opinion is still, that the -- it's very precisely defined, as to what goes into the

outcome, but I'm not sure if it's capturing everything or -- and if it's arbitrary assignment or not. I think that's restating whatever the --

DR. WEINTRAUB: So, the question is, is this -- does it capture the most inclusive target population?

CO-CHAIR CURTIS: Not so much the population, because I think the population is okay. It's the outcome.

DR. WEINTRAUB: Well, but it's still consistent with --

CO-CHAIR ROSENZWEIG: Specifications are consistent, yes.

DR. WEINTRAUB: Yes.

CO-CHAIR ROSENZWEIG: Whether they're accurately represented, I can't tell you.

DR. PALESTRANT: Can I make just one more comment?

MS. TURBYVILLE: Go ahead.

DR. PALESTRANT: Just asking the

group, if you were to -- knowing what you know, which is actually more than what most people would know, when they get a report from this, and they issue a new received report with these numbers, would you know what went into generating that report, in any great depth?

DR. LYNN: Who is that question for?

DR. PALESTRANT: Just for the group, it's my concern. Even having studied this, gone through these a few times, trying to read it, trying to understand how this was all generated.

I'm still not clear, what the metric that would be -- would you get as your answer, what it would actually mean, and therefore, this comes to -- at least the heart of the problem, I mean, we know more than what most users of this will know, and that could be a course for the report.

CO-CHAIR CURTIS: I'd just say

that I have moderate confidence in my ability to understand what exactly, the numbers would mean, all right. I think you could decipher it. I think this -- this is complex. It's been generated over years and years, and clearly, a lot of thought has gone into it.

It's hard for us, in a two hour span of time, to unwind it and make sure we understand.

DR. PALESTRANT: So, you're really endorsing it for general use, correct? I mean, most of us get confused, we're endorsing this as a metric.

CO-CHAIR CURTIS: We're endorsing it as a measure of resource use.

DR. PALESTRANT: Yes, correct.

DR. WEINTRAUB: This is a generic problem, right? I mean, this is a problem with all of these, that we're struggling to figure out what they mean, as fairly sophisticated people, and putting more time into it than most.

If the person who is getting this in a report in a hospital, gets back one of these reports, and I deal with this on the other side, the NCDR, and we developed these reports for people and I'm constantly saying, "These reports are no good. We've got to get better reports."

And I can tell you, people don't understand the reports.

DR. PALESTRANT: Correct, and that was the number, and that's significant, because there is a number for this, and that's the part of the responsibility of -- at least from my perspective, of what we're trying to -- - when we make judgments on these metrics, you have to understand how it will be used, and whether it will be useful, when that data is generated.

And I'm not sure that anybody who gets the score will actually know what it means.

DR. HWONG: So, you know, from my

perspective, coming again, from an analytics company within a health plan, and the health plan, we do support, you know, somebody's physician quality profiling efforts, and generate reports, and although my group isn't, you know, formally involved in the efficiency fact, you know, calculation side, right, you know, for example, within Well Point, you know, a different analytic group actually does use these ETG, you know, the ETG methodologies.

So, what I would -- from my experience, in looking at this, I think whenever sort of scores go out, you have a huge amount of feedback on the quality measures, the process measures, because that's very, in some ways, you know, for physicians and in terms of just training, or whatever, it's just sort of easier to kind of get into and sort of, you know, find issue with, number one.

So, we get a lot of feedback that

way. I think the efficiency side, historically, you know, you do get a lot of frustration about that. You get more of this, you know, sort of -- yes, just not only frustration, because it is difficult to kind of wrap your -- again, sort of wrap your brain around it. It's, you know, sophisticated, right.

But that being said, I think there have been a lot of efforts, certainly from the health plan perspective, you know, when implementing these programs, to try and break it down. I think Ingenix also has, you know, like you said, this website for transparency, to try and explain some of these weights, etcetera.

So, maybe part of this, in terms of, I mean, maybe I'm sort of standing on too much of a soap box, but maybe part of this, in terms of being able to endorse or sort of raise the awareness on a national level about these sort of methodologies is to try and help

-- you know, as they -- as they are used, you know, a lot of these programs, to try and kind of highlight, you know, sort of awareness of it, and I think maybe, you know, we're sort of moving in that direction.

So, the only thing I would say, I recognize, I think you know, it is difficult to understand. I don't think it's impossible to understand. It's going to take a lot of time and a lot of education, but you know, given sort of use of these, and especially sort of the importance, in terms of characterizing sort of resource use, you know, this may be a good step in that direction.

DR. PALESTRANT: That does seem to speak to the black box, and part of what the idea is here, at least from my understanding, is that you'll get a number, or you'll get a score, and then the idea would be in order to contain cost, is that people will make adjustments and then go forward, and hoping their score improves and bring at least,

standard throughout the country to be similar, or at least the same utilization of resources, and that way, we can reduce costs.

If I get this number, as a practicing physician, or as a health group, or as an ACO, what I actually know, how I can change or improve, what I know what this number actually means, so that I can effect changes in my organization, and I would argue that I'm not sure anybody receiving this report would have any notion of what to do about it.

CO-CHAIR CURTIS: But I think you're getting towards the usability issue, which I actually think they do have some better response to, than most.

So, I think for this particular group of votes, we're really just saying, how reliably can they count up the resource use in the population of interest and does it meet that threshold?

So, that is, at least for what

we're voting on right now, and we will get back to the usability, and that's why I brought up at the start of this review, is this really -- do you want us to review this from the perspective of physician profiling, or at the level of the payer or some other population based level, because I think that's again, sort of has different sensitivities.

Well, we did vote on 2b1, so, I think we should go through the reliability and validity.

CO-CHAIR ROSENZWEIG: Are we up to testing and analysis?

CO-CHAIR CURTIS: Yes.

CO-CHAIR ROSENZWEIG: Okay, all right. Okay, so, the reliability testing, they basically had -- used a large health services benchmark database, 25-million covered lives for the calendar year 2009, and 4-million member sample, 7-million member sample used for reliability evaluation.

But this is not specifically for

diabetes, is it?

DR. LYNN: No.

CO-CHAIR ROSENZWEIG: All right.

DR. LYNN: These are all -- these are not just people that have diabetes. It's everybody.

CO-CHAIR ROSENZWEIG: Okay, all right, and okay, and they found that it was internally consistent, and there was a -- and they were also able to look at reliability across HCO's, showing measures of resource use for nine healthcare organizations.

CO-CHAIR CURTIS: What I liked about the description of reliability is that they actually described the internal QI process, which was absent from, I think, the other measure that we've evaluated.

But they have a whole peril process with making sure that it's truly getting to the same result. So, I have a much higher confidence of the internal reliability of this, as opposed to the others.

CO-CHAIR ROSENZWEIG: Yes, okay, and then the validity testing, there again, large number of patient samples for which this review, 7-million member sample in nine healthcare organizations used for reliability assessment, and they were able to process comparisons between ETG and resource utilization software, and got -- DR. WEINTRAUB: So, you developed -- did you -- in one group, you had a delegation, and another group of validation, is that what you did?

CO-CHAIR CURTIS: Microphone.

DR. WEINTRAUB: I'm sorry, did you do a standard delegation and validation study, developing model in one group and testing in another?

DR. LYNN: No, I think what was done here, and I'm not exactly the person who has done it, but I think what was done here is that we looked at -- just looked at metrics across multiple health plans for a consistency.

I don't think it was -- this is different than developing the model of severity. This is not where we develop the model of severity, which we're trying to show that -- some reliability.

We did not do sort of the statistical measure of how close these different health plans were.

DR. WEINTRAUB: So, then this is not truly validity testing, right? All you did is see that you have measures that can be applied in some kind of way, in different populations.

[overlapping voices]

DR. REEDER: I think they're in the beginnings of content and construct validity here.

DR. WEINTRAUB: Well, I mean --

DR. REEDER: In what I'm reading.

DR. WEINTRAUB: I mean, that's sort of a different kind of issue.

What I'm talking about is validity

testing of your model.

CO-CHAIR CURTIS: So, the focus is really not so much on the validity testing of the risk adjustment, but it's really in the reliability of which you can specify the population, and the validity is in the repeatability of the ranges across payers, I think, and I think that is more on --

DR. WEINTRAUB: I think we have to be careful of what we mean.

CO-CHAIR CURTIS: So, I think it's face validity is being supported by that output.

DR. WEINTRAUB: Okay.

CO-CHAIR CURTIS: If I could speak for the measure developer.

DR. LYNN: You're doing great.

(Off mic comments)

DR. REEDER: Just for the record.

DR. WEINTRAUB: Yes, again, I would agree, it looks to me like there is face construct validity, but validity -- there

isn't this formal statistical validation, which could be done.

DR. LYNN: Right, but I don't think we -- we haven't done that.

DR. WEINTRAUB: It's easier.

DR. LYNN: Maybe I'll, you know, get my boss, Dan Dunn, in touch with you.

CO-CHAIR ROSENZWEIG: So, they also describe how they deal with exclusions. They eliminate outliers and they also eliminate a variety of incomplete episodes, okay, and how do you exactly describe an incomplete episode?

DR. LYNN: Again, in the case of diabetes, an incomplete episode is a member who has not been eligible for the year in which there was a diabetes episode.

CO-CHAIR ROSENZWEIG: Okay, and they've also tested this, with respect to resource use between 2006 and 2010. Was diabetes specifically addressed, in this population -- in this particular testing,

analysis of exclusions?

CO-CHAIR CURTIS: You're on page 43, now?

CO-CHAIR ROSENZWEIG: I'm on page, yes, the bottom of page 42 and the top of page 43.

CO-CHAIR CURTIS: Okay.

DR. LYNN: Let me see, I happen to have, I think it's diabetes specifically, but -- this is 9.7.

CO-CHAIR CURTIS: Your microphone.

DR. LYNN: I'm talking to myself. I'm just looking to see if these are different for the two different -- for another one that I have open. It will just take me a second. Sorry for the delay.

CO-CHAIR CURTIS: Maybe we can keep moving forward with the description.

CO-CHAIR ROSENZWEIG: All right, okay, and the analytic method, I think we've kind of discussed this already. I think we've gone through this particular point, as well,

and -- okay, and so, I think we're sort of at the -- I think we're ready to vote on this, on the validity section.

CO-CHAIR CURTIS: So, I'll just the liberty. So, 2a2, reliability, testing demonstrates that the results are repeatable, producing the same result a high proportion of the time, when assessed in the same population, the measure score is precise.

So, in the absence of additional conversation, why don't we go ahead and vote?

CO-CHAIR ROSENZWEIG: Yes, I gave this a high value.

DR. LYNN: Just, that was overall diseases, not just diabetes.

CO-CHAIR ROSENZWEIG: Okay.

CO-CHAIR CURTIS: That's seven high.

DR. WEINTRAUB: Have we lost someone?

CO-CHAIR CURTIS: Brenda.

MS. TURBYVILLE: Brenda Marie is

next door.

CO-CHAIR ROSENZWEIG: The person on the phone, does he vote, too?

CO-CHAIR CURTIS: So, 2b2, validity testing demonstrates the measure data elements are correct, and the measure score correctly reflects of care, resources provided.

CO-CHAIR ROSENZWEIG: Yes, it looked to me like it was -- you know, internally, it certainly seemed like they were measuring costs and comparing them between groups.

So, here, again, gave it a high rating.

CO-CHAIR CURTIS: I gave it a moderate, just based on that heart failure/diabetes example that -- again, I don't know if it's capturing the true total costs, but that's my take.

DR. WEINTRAUB: Well, I mean, we haven't seen formal evidence of discrimination

and we certainly haven't seen calibration, unless I'm missing something. I can't give more than a moderate.

MS. TURBYVILLE: One more vote?

CO-CHAIR ROSENZWEIG: Okay, so the next one is 2b3, exclusions are supported by the clinical evidence, otherwise, they are supported by evidence of sufficient frequency of occurrence, so that results are distorted, within -- with the exclusion.

You know, there is a discussion of exclusions here, but a lot of it is very much based to whether or not it fits within the grouping.

So, I didn't give it a high rating. I gave probably either moderate -- I started -- when I originally reviewed it, I thought it was low, but I'll move it up to moderate, from my recommendation.

CO-CHAIR CURTIS: Waiting on one, and seven moderate, okay.

CO-CHAIR ROSENZWEIG: Okay, and

the risk adjustment strategy, here, we're getting into a certain amount of data analysis that does demonstrate that the methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically, clinically meaningful differences in performance.

See, this is sort of an area in which Connie comes from one end of the spectrum, in which she's looking at physicians and they're getting scored okay, and from their perspective, there is -- they're getting data that seems internally consistent and is reliable.

I don't know, as a physician being judged, I tend to be a little more skeptical about these scores, that when I get them, and I think -- I agree, that they really put a lot work into this and that I'm not being able to totally judge methodology, from my own perspective.

So, I can't give it a one. I

would either give it a moderate or -- probably a moderate score, or perhaps a low -- I mean, because frankly, physicians are much more -- when they get scoring related to quality care, it's much -- it appears much more transparent to them, because it shows the number of -- their percentage of patients that get A1C's done and percent, and what their average A1C is and so forth.

But here, we're getting basically, a score that gives you sort of an estimate of your costs related to other physicians costs in diabetes, and it does -- and absent of the clinical -- and of course, we're not expecting this from the developer, but absent some quality assessment, that maybe your costs were deservedly so, or not deservedly so, it has a certain -- less amount of meaning, to a certain extent.

DR. LYNN: I just want to make one comment, and that is about how, you know, it's -- it's hard for this stuff to be actionable,

and it is hard for this stuff to be actionable, but one of the things that we've tried to do is include these other measures, not just the cost, but you know, number of ER visits per episode and hospitalizations per episode, to try to help get at some of these drivers.

You know, there is a lot more that needs to be done, but you know, we are trying to provide some drivers of cost.

DR. HWONG: And as far as like 2b4, so, Jamie, I think your points are well taken.

You know, for this, is this really that sort of risk adjustment methodology? Are we ranking that, 2b4, or --

CO-CHAIR ROSENZWEIG: This is 2b5.

DR. HWONG: I'm sorry.

MS. TURBYVILLE: No, we're on 2b4.

DR. HWONG: We are on 2b4?

CO-CHAIR ROSENZWEIG: I'm sorry.

DR. HWONG: I'm sorry, yes, I was

looking at the screen and getting kind of confused. Yes, we're voting on risk adjustment or -- and then I was thinking about -- I was thinking he was talking about the next sort of topic.

CO-CHAIR ROSENZWEIG: All right, I will address that, I apologize.

DR. HWONG: So, I'll hold that thought.

CO-CHAIR ROSENZWEIG: I skipped one of the measures. Okay, so, for outcome measures and other measures, when indicated, an evidence based risk adjustment strategy, risk models as specified and is based on patient clinical factors that influence the measured outcome, but not related to disparities of care.

I would give this, yes, a high rating. I'm sorry, I apologize to the rest.

DR. WEINTRAUB: Really, I mean, this has problems here. First of all, you automatically distinguish between

complications and co-morbidity, and then I think that we haven't seen enough, in terms of specification of the model.

So, I think was going to give it a low. You could talk me into a moderate, but you'll have troubling telling me it's high.

CO-CHAIR CURTIS: I share those concerns, particularly, you know, without a very strong explanation or rationale for why you're identifying -- can't distinguish between complications and co-morbidities, it flies in the face of most measures.

It might be still as accurate, I don't know, but it certainly violates our principles. So, maybe too strong.

DR. WEINTRAUB: I mean, you know, talk about colinearity, I mean, complications predict complications. If you have heart failure, you have heart failure. You absolutely cannot make sense of putting into a model, complications to predict the outcome.

DR. LYNN: I think it's a

different question, slightly. You know, are you going to hold physicians responsible for the greater difficulty of caring for diabetes, because they have something else that's going on, that whether it occurred before or after, is it "their fault"?

And again, the model does not include the cost of caring for the heart failure. It includes only the cost of caring for the diabetes.

DR. WEINTRAUB: All right, so, if you take it out and you don't count it, because that admission for the studies have nothing to do with diabetes, no problem. They're not getting dinged on it.

CO-CHAIR ROSENZWEIG: All right, you know, considering all the complicated issues related to co-morbidities, yes, I'll lower my recommendation from high to moderate, yes.

CO-CHAIR CURTIS: I think the other piece, though, is that if we are pairing

this with quality measures in the future, you can't really have two different approaches to risk adjustment, and no one is ever going to modify, to identify complications as co-morbidities.

DR. WEINTRAUB: The last thing you want to do is have a measure out there with complications, as part of the model.

I mean, we criticize from here to eternity.

CO-CHAIR CURTIS: So, let's go ahead and vote.

CO-CHAIR ROSENZWEIG: Okay, so, people were more skeptical of this than I was, okay.

And then the next one, data analysis -- I'm sorry, this is 2b5, data analysis demonstrates that methods for scoring an analysis of the specified measure allowed for identification of clinically significant and practically clinically meaningful differences in performance.

Well, you know, since you're looking at these differences on so many different levels, it's hard to say.

I suspect that probably the data analysis for large groups could really show differences in performance, within the spectrum of what is being measured here.

I wonder about the number 30, that was given to us, with respect to looking at individual physicians, especially since many physicians may not have that many patients with diabetes in their practice, and so, I would be a little skeptical about that one and I'd probably give it a low reading.

CO-CHAIR CURTIS: Let's go ahead and vote then.

CO-CHAIR ROSENZWEIG: Two moderate and -- excuse me, four moderate and three low.

MS. TURBYVILLE: Could I make sure I capture the rationale, here? I was still working on the risk adjustment rationale, quickly, so, I didn't hear what Jamie said.

So, is it -- was is, going back to the issue of the physicians, or -- I completely missed what was said.

CO-CHAIR CURTIS: It's insufficient evidence of a threshold number, or that the -- that the results are clinically useful at the end of the day, at the level of the physician, less so for higher levels.

MS. TURBYVILLE: So, unlike the other measures, because it's recommending the physician level of analysis?

CO-CHAIR ROSENZWEIG: I'm sorry?

DR. LYNN: This is kind of an interesting issue. I mean, you know, if you give it a different rating, if you use the group or -- MS. TURBYVILLE: Right, so, I just want to make sure I'm capturing that.

So, is it -- so, unlike the other measures, this one, in looking at physician profiling, it's a concern that it wouldn't be actionable by an individual physician, or is

it also about something else that I've missed?

CO-CHAIR ROSENZWEIG: Yes, I have a certain amount of skepticism about the individual profiling of individual physicians.

MS. TURBYVILLE: Okay.

CO-CHAIR ROSENZWEIG: With respect to this, because of the number of patients with diabetes within a specific panel that physicians tend to have, and so forth.

I have less skepticism with respect to analyzing the data and like NCQA does, to look at plans, since they -- it's likely that they're probably -- those are more consistent.

MS. TURBYVILLE: Okay.

CO-CHAIR CURTIS: But looking at -- there needs to be consistency across, right, because the ABMS measures also specific the level of the physician, and I don't know if we have a similar low range.

MS. TURBYVILLE: But I'm wondering if there is a little change in the tone of the

--

CO-CHAIR CURTIS: So, we can look at that, when we get to the comparisons.

MS. TURBYVILLE: We'll revisit, as long as it's captured here.

CO-CHAIR CURTIS: Well, do we want to re-vote on that, with that consideration, because I don't think it's different.

CO-CHAIR ROSENZWEIG: Well, but the ABMS also used a very widely recognized system that's been in place -- you know, that's, at least from my perspective, is a little more not proprietary and is more transparent.

So, I don't know, I can't remember what we actually voted on for the ABMS value for this particular measure, but --

DR. WEINTRAUB: Can we look at it?

CO-CHAIR CURTIS: It will be from this morning, for the diabetes measure, specifically.

CO-CHAIR ROSENZWEIG: The diabetes

measure, specifically, yes.

CO-CHAIR CURTIS: Not the NCQA, just the --

MS. WILBON: The diabetes measure is not --

MS. TURBYVILLE: Yes, it's not. This is --

CO-CHAIR CURTIS: It's not in there?

MS. TURBYVILLE: This software tool does not allow us to easily summarize after it's run, sorry.

MS. WILBON: I can read it out loud, hold on one second.

MS. TURBYVILLE: Okay.

CO-CHAIR ROSENZWEIG: Okay, then if multiple data sources are specified, and there is demonstration that they produce comparable results, I don't think that was addressed here, specifically.

CO-CHAIR CURTIS: Right.

CO-CHAIR ROSENZWEIG: It's all

from the -- yes, yes, yes.

CO-CHAIR CURTIS: So, I summarily move that we dismiss 2c.

CO-CHAIR ROSENZWEIG: That being non-applicable.

CO-CHAIR CURTIS: Yes.

CO-CHAIR ROSENZWEIG: I suppose.

CO-CHAIR CURTIS: Right.

CO-CHAIR ROSENZWEIG: Okay, and then so, I think we're onto usability.

MS. WILBON: So, just a quick follow up to the request before.

I think you wanted your ratings on the ABMS diabetes measure.

CO-CHAIR ROSENZWEIG: Yes.

MS. WILBON: Specifically on 2b5?

CO-CHAIR ROSENZWEIG: Yes.

MS. WILBON: Around statistical meaningful differences. Six high and two moderate.

CO-CHAIR CURTIS: Unless we can --

DR. HWONG: You know, I mean, I

think I wonder if it's there, with ABMS. Remember, they sort of specified very clearly in sort of clinical terms, like what services are going to be included, whereas, maybe what's getting -- maybe -- potentially, there was a difference, like, you know, here, it's still, there is that interplay with different episodes and kind of what ultimately ends up inside.

You know, may be less interpretable to a physician at the end, even if I were to hypothesize what might be some of that difference.

That being said, I do recognize that the Ingenix developer, especially when you stratify -- not stratify, but like, when you -- well, that's sort of strata, but like, pull out, in terms of the different resource categories, right, but you know, I mean, I think that does go part of the way to address, you know, it being a meaningful, something that potentially could be meaningful, in terms

of someone's practice or a group practice.

CO-CHAIR CURTIS: And I think it also does get into the risk adjustment methodology, as Jamie said, is different and there are more significant questions about that risk adjustment methodology, you know, years prior or within the year, and that may be affecting people's votes, certainly.

CO-CHAIR ROSENZWEIG: The other issue is that the other group also specified that they would only use it to compare physicians when there was statistically significant differences, as least as I recall. That's not the case?

CO-CHAIR CURTIS: I think that's how they specify here. I don't think that they're applying it differently.

CO-CHAIR ROSENZWEIG: They're just saying 30.

CO-CHAIR CURTIS: No, no.

MS. TURBYVILLE: That's just an example.

CO-CHAIR ROSENZWEIG: That's just an example?

CO-CHAIR CURTIS: No, that's just an arbitrary number.

CO-CHAIR ROSENZWEIG: Okay, all right.

MS. TURBYVILLE: So, we'll just revisit that when we get more feedback from all the measure developers, including their R-squared for their risk adjustment method.

CO-CHAIR CURTIS: Okay.

DR. WEINTRAUB: Ask him to give calibration, as well, not just discrimination.

MS. FANTA: We've got it down.

CO-CHAIR CURTIS: So, then we're on usability?

CO-CHAIR ROSENZWEIG: Yes. Okay, we're on page 45, and -- I think that -- they certainly have reports that seem to be readable and logical, and seem to be reasonably current, and they compare -- but what they're doing, at least with respect to

their discussion of this, they're really talking about comparing healthcare organization one versus various others, and they're not specifically talking about diabetes.

CO-CHAIR CURTIS: Do you want to comment on that, Tom?

DR. LYNN: Let me look at it, before I do, because -- is this still --

CO-CHAIR ROSENZWEIG: I'm talking about U11 and U12.

CO-CHAIR CURTIS: Page 45 and 46, I think is the predominant, and -- the other piece of this is that they talk about how the payers are using it. They're not talking necessarily, about individual providers, but they don't state that they payers are not looking at the individual providers.

So, you know, it's a little bit of a black box, but this is the first time that we've really gotten to usability, to a large extent.

DR. HWONG: Right, and the only thing I would add to that, in terms of coming just from one payer, you know, especially where the symmetry, in an ETG product that's used, so, as far as where it's used for like, community, you know, public reporting, there are listings, in terms of the provider directory, you know, there is program that looks at sort of overall cost and quality rankings and so, that is the driver of that.

So, in some ways, like, physicians, the community at large, actually sees that. So, if this category is really about like, you know, are these performance results sort of made public? Are they seen? Are they used in programs? I know of at least, like, personally, like one example where that is being done.

DR. LYNN: This is not specifically talking to that, overall.

DR. HWONG: Right.

CO-CHAIR CURTIS: Is there an

overall Ingenix measure under evaluation in this process, one that rolls up all the individuals?

DR. LYNN: We had to -- there is actually a bug on my -- you guys are bugging my stuff.

There is actually -- we had a measure, but there were complications that were not technical. They were more -- other complications that we had to take, that went down.

MS. CLARK: I have a question about this public reporting. I mean, it says public at large. I mean, does this really mean that it needs to be -- it's widely available to the public, right?

I mean, all of these uses are really internally within the health plans, as the examples. So, is that considered public?

CO-CHAIR CURTIS: What Connie just referred to is private.

DR. HWONG: Granted, it's not in

the application, but like, I can just from my experience, right, and just one single example --

MS. CLARK: Okay.

DR. HWONG: -- but the rankings for physicians are placed on the website for the provider directory.

So, you can see the results, and then there is information about the program, as to how those scores, or how those symbols, the blue ribbon, gold star kind of stuff, right, you know.

But, you know, it's certainly out there, that someone, you know, a lay person can go to the website, take a look, right, and see, you know, look up their physician, that sort of thing.

MS. CLARK: But that's -- is it the same report, type of reports that are here? I mean, it's a different report?

DR. HWONG: Right, yes, I hear you. So, for the one example I'm giving in that regard, that -- the full report isn't

available to patients at large, that way.

That being said, in the other programs that we do, in terms of, it is available to the physicians, to the physicians, in terms of their individual reports.

So, yes, it may not get at, Mary Ann, you know, at this sort of wide spread public, you know, dissemination, maybe, but there is some --

MS. CLARK: Was that the intent, though?

MS. TURBYVILLE: That's a conversation that continues to occur at the CSAC level, our committee for standards setting that oversees all of the Steering Committee work, and there is discussion about, you know, is it public at large? Does it have to be the General Joe, and Heidi might want to add to it, but I'm not sure that that is, you know, 100 percent where we are right now.

MS. BOSSLEY: I think this is

evolving, and in fact, the Board will be looking at a recommendation, and we have a task force that specifically looks at, when we talk about usability, what does that mean, the first time you see that measure, what kind of use are we looking for, and the usefulness, because there is two pieces to it, and then at maintenance, what are you going to?

So, hopefully, as you're going through this, you will have a group that is advising, and it will probably help you refine your criteria on that, as well. But this is one of the more loose ones we have, right now.

CO-CHAIR ROSENZWEIG: Okay, so, with respect to demonstration of usability, you have used this in public reporting, in a variety of settings, at least you say here, but they ask, at least the -- NQF asked for the names of the programs, the locations and web URL's and you're just basically telling us HC-05, HC-06, and so forth.

Is there -- is that proprietary

information that you can't divulge?

DR. LYNN: I actually don't know the answer to that question. I can find out, and get back to you about that.

CO-CHAIR ROSENZWEIG: Are there specific -- is there any data that -- the use of these, you know, this reporting has actually resulted in quality improvement or in cost reduction?

CO-CHAIR CURTIS: Well, that really gets to the effectiveness, which I don't think was the criteria.

The criteria is whether or not it's being used by --

CO-CHAIR ROSENZWEIG: Okay.

CO-CHAIR CURTIS: -- people for quality, attempts to improve quality, and I think that they do meet that threshold.

CO-CHAIR ROSENZWEIG: Okay, it's been used?

CO-CHAIR CURTIS: It's being used by --

CO-CHAIR ROSENZWEIG: Not specifically for diabetes, at least as a diabetes stand alone.

CO-CHAIR CURTIS: Right.

DR. WEINTRAUB: That's important. There are examples here, are things entirely different, Caesarean section, for instance.

CO-CHAIR CURTIS: But that gets back to this sort of artificial, the construction of the overall ETG methodology and to its component parts.

So, you know, they don't ever report out diabetes without the larger context. So, it might be impossible.

DR. LYNN: It's not that we don't ever do it, but that is not what is usually done.

CO-CHAIR CURTIS: Well, right.

CO-CHAIR ROSENZWEIG: But the plan is, in the future, to. That's why you're coming to us with this measure, right?

DR. LYNN: Honest, we see this as

a part of a bigger effort. You know, we assume that this is going to continue to go and there are going to be more diseases and more diseases added, and we recognize that you need -- you know, part of increasing the end here is increasing the number of diseases that are sort of certified, and I know that's time consuming and difficult.

DR. MARWICK: I think if that's the goal, then dealing with the ambiguity about how people get allocated from bucket to bucket, that's something I personally would feel much more comfortable about, and still feel some disquiet about that.

CO-CHAIR ROSENZWEIG: Okay, testing of interpretability, this is interesting. I mean, the interpretability has been looked at by the medical advisory board of Ingenix and also, the user.

Could you explain who the user forums are? I mean, have you done any testing, like, actually sent out

questionnaires to the physicians who are being rated by this?

DR. LYNN: We probably have not sent out questionnaires -- You know, being that we're not the organization that actually measures the physicians, we haven't gone out to the physicians.

These are probably -- this is basically input from the intermediary, who are using the methodology.

So, our users are, you know, health plans, large provider organizations and groups like that, that use this methodology.

CO-CHAIR ROSENZWEIG: Okay, you're not giving us any specific details on the data that's been reported to you. You're just saying that they were --

CO-CHAIR CURTIS: Is this in contrast to the NCQA diabetes measure? I know we talked about how they -- I think they specified that they have focus groups and commented on the usability of the measures and

things like that.

But again, I don't know if it was that specific to say, this company said that this was useful for these reasons.

But I agree, that as a product, as it matures in this arena, it might be useful to do that type of testing in the future.

CO-CHAIR ROSENZWEIG: Okay, and resource use data and result can be decomposed for transparency and understanding.

DR. WEINTRAUB: Can be.

CO-CHAIR CURTIS: Actually, in the sense, I feel like I'm on the promotion committee now, but if you couple it with the ER visits, with the individual service lines, I think there is that potential for enhancing the interpret-ability and the decomposition of the total costs, to the component elements.

Again, I don't know if that's been adequately demonstrated for our group, but --

CO-CHAIR ROSENZWEIG: I wasn't -- I mean, it -- I suppose it could be, whatever

decomposing means.

Okay, and if this measure has either the --

CO-CHAIR CURTIS: I worry about Brenda, it might still --

CO-CHAIR ROSENZWEIG: Has either the same measure focus or same target population as NQF endorsed measures. Are these measure specifications completely harmonized?

DR. HWONG: I think we're leaving that off the table right now, the harmonization.

CO-CHAIR ROSENZWEIG: I think that's off the table, yes, I wouldn't -- I had it as insufficient or not applicable. So, let's make it not applicable, so, we don't have to -- okay.

All right, do we want to go over these now, or we can just continue through the --

CO-CHAIR CURTIS: I think I'd stop

at usability --

CO-CHAIR ROSENZWEIG: All right.

CO-CHAIR CURTIS: -- and then maybe go through it, yes.

CO-CHAIR ROSENZWEIG: All right, so, for 3a?

DR. HWONG: Yes.

CO-CHAIR ROSENZWEIG: This is for the use and quality improvement. They certainly listed a series of organizations. They haven't specified what they are, and haven't specifically indicated that there is -- that there is specific benefit from them.

But it certainly seems like they are probably pretty usable, from my perspective. I had the sense that they would be usable. So, I gave this a moderate.

CO-CHAIR CURTIS: So, I don't believe this is for quality. This is for the public reporting aspect.

CO-CHAIR ROSENZWEIG: This is for the public reporting, as opposed to whether or

not it actually -- whether or not it -- yes, as you -- I'm not -- there is no -- there is not a lot of evidence that it specifically produced better quality, but there is evidence that it was usable, okay.

DR. WEINTRAUB: In that sense, it benefitted the public?

DR. HWONG: I got a sense that this was more about just, is it out there, right? Is it publically available? Is it somewhere? It's not just hidden in a closet somewhere for some private purpose, but that, you know, that there is some opportunity, potentially for some feedback on the actual, you know, structure and method.

CO-CHAIR CURTIS: And I think we're voting on our understanding of overall Ingenix measure, and not necessarily the specific line within it.

I think you have to have that, because it's --

DR. PALESTRANT: I think my

concern is publically reported occurrence -- it was really isn't to any great extent, and could it be in the future? Sure, but it's not, and then does the report look like, is also a big question.

If you publically report it, then it has to have the details of how this measure was devised, and not -- and so, is the company then prepared to sort of -- release a lot of data to the public about how they came to these numbers, and I don't think we have the answers for those questions, right now at all.

CO-CHAIR CURTIS: So, it seems like it might default back to non-applicable, or --

DR. PALESTRANT: Well, no, no, no, well, first of all I think it's either high, not low --

CO-CHAIR CURTIS: Or insufficient, one of the two.

CO-CHAIR ROSENZWEIG: Your point is well taken, because none of these are

diabetes specific, so, they're not really addressing the specific measure.

DR. HWONG: I guess, I sort of look at it, again, you know, understanding this carved out from the overall methodology, and then I look at the sort of sub-sub-criteria, right, the -- is it currently in use? You know, is it used in public reporting initiative, which I am hearing from everybody, that, you know, I feel like less confident about that.

Is it used in quality improvement efforts, that I see, that that is, you know, based on sort of the responses, you know, placed in the application, and used for other accountability functions, as well, you know, in terms of the QI and if there was sort of accountability, you know, at a physician level.

So, yes, I guess when I look at the actual, sort of the sub-sub-criteria, I think there are, you know, some aspects, you

know, maybe not all, right, but some aspects that might be fulfilled.

CO-CHAIR ROSENZWEIG: As I said, it's not the diabetes measure. It's the ETG methodology.

CO-CHAIR CURTIS: I just think it's hard to separate it out, because that is the evidence that they've given us.

MS. TURBYVILLE: Right.

CO-CHAIR CURTIS: If it's evidence of the diabetes, it's got to be insufficient, by definition, right, because they haven't provided that level.

So, I guess we could vote, either way, we could choose -- I mean, we should probably be agreeing as a group, as to which -- you know, and I'm comfortable voting on the overall, as long as that's in the annotation to the Steering Committee, that that is how we took this.

Sally, so, do you want to -- do you have a thought, as to which direction we

should go?

MS. TURBYVILLE: Well, we are charged with evaluating this as an independent measure. The current text and the background of the ETG system is clearly, critical to understanding the measure, and how it might be used.

But the endorsement process, as we have is structured, and we had similar comment to ABMS, when they were stating paired measures, really is an independent evaluation.

CO-CHAIR CURTIS: Okay, so, I think that's adequate guidance. So, okay, let's vote. One moderate and one low, four insufficient.

CO-CHAIR ROSENZWEIG: Okay.

CO-CHAIR CURTIS: So, for 3b, this is where the results are meaningful, understandable and useful to the intended audience.

CO-CHAIR ROSENZWEIG: So, they showed a mock-up of what the report looks

like. They've used -- for the purposes of this, I don't think it'd need diabetes specificity as much, at least that's the way I would interpret it.

They have had sort of a -- sort of reviewed by their medical -- their own medical advisory board, an Ingenix user forum, but it hasn't really be tested for interpretability by -- at least, they haven't given us evidence that it's been tested for interpretability by outside groups.

CO-CHAIR CURTIS: Okay, so, should we go ahead and vote on that?

Okay, four moderate, one low and one -- two insufficient, and 3c, data and result details and maintains such that the resource use measure, including construction logic could be decomposed to facilitate transparency.

So, I think we talked about how one could, but it takes an awful lot of effort.

One high, two moderate and four low.

CO-CHAIR ROSENZWEIG: And then the fourth one--

CO-CHAIR CURTIS: Harmonization, we can --

CO-CHAIR ROSENZWEIG: Harmonization, we can punt on.

CO-CHAIR CURTIS: Yes.

CO-CHAIR ROSENZWEIG: Okay.

MS. TURBYVILLE: Can I take a step back for the rationale on 3c, because I think there are a lot of moderate and lows, if I remember correctly, the one that we just did.

CO-CHAIR ROSENZWEIG: Okay.

MS. TURBYVILLE: Okay, so, I just want to make sure, as we capture the rationale, so, unlike for example, the NCQA measure, there ABMS measure, the challenges for the user or the person getting measured, in decomposing, is higher because -- and that's not a challenge, I just want to make

sure or not, or -- so, is it the specification is not having enough of it, to allow for that? Is it the complexity? Is it the relationship to the other ETG's? I just want to make sure I -- or was it all of those things?

CO-CHAIR CURTIS: You're looking at me, but I voted moderate--

MS. TURBYVILLE: No, I'm looking at everybody else. I'm sorry, Jeptha, just was looking so intently and thoughtfully.

So, just to -- so, anyway --

DR. LYNN: Because, I mean, it's rudimentary, but this measure does provide some forays into what drives costs, ER counts, hospitalization, things like that.

MS. TURBYVILLE: Okay, okay. So, any thoughts on this?

CO-CHAIR CURTIS: I think it might be just sort of varying, like, depending on what time of day, how we think about what decomposition means --

MS. TURBYVILLE: No, that's fair.

CO-CHAIR CURTIS: -- and so, I think it probably, I would feed it back, rather than re-vote, I would say that maybe we need additional guidance from the Steering Committee --

MS. TURBYVILLE: Yes.

CO-CHAIR CURTIS: -- et cetera, as to really, what this particular element means.

MS. TURBYVILLE: And what I'm kind of fishing for here is, it may be more input, also, from the measure developer, as we prepare for--

CO-CHAIR ROSENZWEIG: I think it's possible, yes, I think it's possible that it could be decomposed --

MS. TURBYVILLE: Okay.

CO-CHAIR ROSENZWEIG: -- for more transparency and understanding, that -- at least at the level that we've been evaluating at the present. It's difficult for us to assess the extent of that.

CO-CHAIR CURTIS: And I think

that's true for all of the -- well, the two measures that we've gone through, to this stage.

MS. TURBYVILLE: Right, okay.

CO-CHAIR CURTIS: So, Jamie, do you want to go through feasibility?

CO-CHAIR ROSENZWEIG: Did we do 3d?

CO-CHAIR CURTIS: I think we're skipping, we're punting.

CO-CHAIR ROSENZWEIG: Okay, 3d, okay, so, basically, feasibility, the 4a, this is -- this measure is -- the data elements are generated as a byproduct of care processes. Certainly, that is the case and it is generated and used by healthcare personnel, including a whole variety of specific information.

I didn't know that blood pressure was specifically being measured as a part of this. Lab values and medical conditions --

CO-CHAIR CURTIS: I think that is

a drop-down box from NQF, maybe, or -- because I think that has been on the other ones.

MS. TURBYVILLE: Yes.

DR. WEINTRAUB: That is a very important point, blood pressure is not going to be found in--

CO-CHAIR ROSENZWEIG: I don't think you're going to find that in claims data.

DR. HWONG: That is just generic language.

CO-CHAIR ROSENZWEIG: Right, unless you're using CPT categories, category two codes. Oh, so, this is generic?

CO-CHAIR CURTIS: Well, I think it's -- the blood pressure --

MS. BOSSLEY: Generic, EG, the whole statement is a generic check box that they check.

CO-CHAIR CURTIS: Right.

DR. HWONG: Right, the language is, yes.

CO-CHAIR CURTIS: So, the one that is relevant here is medical conditions, as assessed by administrative data.

MS. TURBYVILLE: Exactly.

CO-CHAIR ROSENZWEIG: Well, I would say certainly, that is the case, yes, and electronic sources, yes, all data elements that are not from electronic sources -- are you using anything other than electronic sources? Okay.

DR. LYNN: No.

CO-CHAIR ROSENZWEIG: That is what I thought. So, 4b, and then susceptibility to inaccuracies, errors, and unintended consequences, wow, I mean, they mention small sample size here. Is that also something that's generated -- is that something that was --

MS. TURBYVILLE: Can you scroll down to that?

CO-CHAIR ROSENZWEIG: Okay, so, here again, now, I guess this is -- you know,

I think the issue is largely the inaccuracies, errors and unintended consequences, my worry would be in small sample size per physician, generating a score that could be used for tiering of physicians, that might not necessarily be appropriate.

So, we have to worry about that particular issue. I have less of a concern about this being used to compare individual plans or large provider groups, so to speak, with respect to their costs, okay.

So, that would be the issue related to that. So, you want a reasonably sized peer group, which is what they mention in here as being a factor, as well, to be able to do that, and to a certain extent, the company, Ingenix, understands this better than -- how to evaluate this, better than anyone else.

But it is an issue that does come up. There is a tremendous amount of concern and anger and frustration in the medical

community about physician tiering based upon costs of care, which is being implemented, and to physicians, it looks like a black box, okay, and it affects whether or not physician's co-pays are changed, at least in Massachusetts, it affects whether or not they're listed on lists as preferred physicians for individual plans, and it's usually based upon two criteria, quality of care, which is much more transparent, and then some sort of a score of their costs, compared with the population as a whole.

And to a physician, this often looks like a black box, so to speak, and their lawsuit -- I know, was -- there is a big lawsuit in this, in the State of New York, and there was a -- in Massachusetts, this has gone to the -- it's still being subject of -- I think it's the Board of Medicine in Massachusetts, with respect to this issue, there a lawsuit involved with that, on behalf of the Mass Medical Association.

So, these are complex issues for obvious reasons, and so forth, and so -- am I going on too long?

DR. HWONG: Jamie, only one thing, I might want to comment on.

I agree with you, like there are these complications, in terms of limit supply in some of these settings.

But I get the sense that, you know, it's not that Ingenix creates this for that one expressed purpose, right. I mean, in terms of how it's ultimately implemented, I think there is sort of -- you can have sort of different business rules, different, you know, programs and how you want to use it.

So, I just sort of want to make sure that we were evaluating this, that it's less on, you know, sort of like downstream specific, you know, some implementation, some kind of program in a way, but much about, is this able to kind of discern, for whatever you do with it in the end, you know, I mean, it

has nothing to do with, you know, score -- you know, how you want to do sort of tiered co-pays or what not.

But, just, you know, does it have the ability to kind of, you know, make, you know, allow you to discern between sort of costs that are generated from one physician to another?

CO-CHAIR ROSENZWEIG: Well, I think, you know, I agree with you. I'm not suggesting that Ingenix would -- this is not being used by Ingenix, for tiering. It's be used by the plans, and the plans have -- and in fact, there have been NQF specified measures, quality measures, that have been misused by plans, as well. Not NQF, but NCQA HEDIS measures that have been used by plans in wrong ways as well.

So, I think it's not the -- the question is whether or not it's susceptible to inaccuracies and errors or unintended consequences. That's my concern here, and I

would think that there is this susceptibility.

CO-CHAIR CURTIS: I think that's overall, a good issue that you raise and one that we should -- we discussed at the level of Steering Committee, it's how specific we need to be and how these measures could be used in isolation, as a resource use measure, as opposed to one that's getting more at value.

So, you know, I think we can discuss it further at that level.

CO-CHAIR ROSENZWEIG: I think it's something that NQF as an organization probably needs to think about, as it produces these measures.

CO-CHAIR CURTIS: Also, it's a --

CO-CHAIR ROSENZWEIG: And be able to specify how they might be used, or the limitations.

CO-CHAIR CURTIS: Right, but it's a very thin ice for them, from their other set of consumers, which are the people that are developing the measures and want to -- defer

them to use the measures.

So, it has to do with what is the scope of the purview, and I know there is back and forth at very high levels.

CO-CHAIR ROSENZWEIG: Yes, I understand.

DR. PALESTRANT: But the crux of the matter there is that it's relying -- I endorse the interest endorsement, it's a big deal for many of these - the providers of these measures.

And so, if they can -- they can then at least market their measures, for doing this work, and so, what I could see, not just with Ingenix, and it's not that -- it may actually be unfair, I've got to get to that in a second.

But basically, what's been applied is, that in many of these metrics that we've looked at, they've never actually been used in the past, for the purpose for which they're now being evaluated.

And it's kind of difficult then, to endorse them, if there is no track record, and you don't want to endorse something that then is going to have widespread use.

CO-CHAIR CURTIS: I mean, I disagree that there is no track record. There is no track record, necessarily for public reporting, but there is a track record for its use in the estimation of cost.

DR. PALESTRANT: I would absolutely disagree. I mean, for each of these metrics, if you realize whether it be diabetes, coronary artery disease, they're applying these metrics, these methods for analysis, to what's being asked, and very few of them have been able to give us long substantial track record of data, and you say that Ingenix hasn't, and from what I've been seeing of the other ones there isn't a lot of track record.

DR. WEINTRAUB: I agree with that, absolutely, completely, where, you know,

we're -- we're breaking new ground.

The question is, where are we with that, and how do we move forward?

DR. PALESTRANT: Well, I think you may be putting the cart before the horse, at least it seems to me.

DR. WEINTRAUB: Maybe so.

CO-CHAIR CURTIS: Perhaps we should go ahead and vote on feasibility. I mean, don't want to curtail the conversation, but I think it might be --

CO-CHAIR ROSENZWEIG: Sure.

CO-CHAIR CURTIS: It's not beyond the -- it's a very broad question, and it's beyond, I think, the scope of this individual TAP, and I think the message can be sent upstairs to the Steering Committee, but probably even higher.

Again, that there is discomfort within the TAP, as to, you know, are we accountable, at the end of the day, for how these measures are being applied to our peers,

for the clinicians and otherwise?

CO-CHAIR ROSENZWEIG: But just to mention, 4d, I do think that they have had a lot of experience in the use of, not this measure, but other measures in a variety of situations.

They have a lot of clients and they've used them for that purpose. So, they have a strategy for data collection.

DR. PALESTRANT: I understand that, and I don't want to belabor this, but and basically, we'll get to it. I reviewed the stroke measure, and it's just quite clear, that this is not being used, and you look at some of the examples that they give, using the databases, and there is some problems with the examples that they give.

That gives me pause to think, "Can this actually be extracted to all these different metrics that we're asking them to do?"

You know, from that point, are we

going to be able to do that when we review that section?

CO-CHAIR ROSENZWEIG: Okay, let's vote on the feasibility.

Okay, the first one is routinely generated and used during care delivery. They're not -- okay, all right.

CO-CHAIR CURTIS: I think we've clarified this measure. It's administrative and routinely generated.

CO-CHAIR ROSENZWEIG: Okay.

CO-CHAIR CURTIS: And then for two, that it's available in electronic format.

CO-CHAIR ROSENZWEIG: Yes, yes.

CO-CHAIR CURTIS: I think there is going to be sort of a pro-forma.

CO-CHAIR ROSENZWEIG: Available in electronic format. The third one is susceptibility to inaccuracies, errors, and unintended consequences.

So, I assume a high score means that it's not susceptible and a low score

means that it is susceptible.

DR. WEINTRAUB: Or it can be minimized.

MS. TURBYVILLE: Right, there is the 'or detected'.

CO-CHAIR ROSENZWEIG: Okay, or it can be monitored, okay, all right.

MS. TURBYVILLE: In this case, high means it is --

CO-CHAIR ROSENZWEIG: It is susceptible to inaccuracies.

MS. TURBYVILLE: That it is not, okay.

CO-CHAIR ROSENZWEIG: Yes.

CO-CHAIR CURTIS: I hope that is how people have been voting all along. We probably should have clarified that yesterday morning.

CO-CHAIR ROSENZWEIG: Do we want to do this again, so that -- everyone clear that they voted the right way on this?

DR. WEINTRAUB: Yes.

CO-CHAIR ROSENZWEIG: Okay, good, okay. Two high, two moderate and three low, very evenly divided, okay.

Then susceptibility -- then the last one is the data collection strategy measure is in use.

Four high, two moderate and one low, all right, thank you.

DR. LYNN: Thank you.

CO-CHAIR ROSENZWEIG: Thank you for your help.

DR. WEINTRAUB: I'd like to bring up a general issue that came up, it's sort of been percolating in my mind. I don't think it has to do with anyone, but --

In doing this kind of modeling, what is an acceptable R-squared?

Now, what kind of R-square do you expect? I'll tell you what the R-square is. You know, the R-square, as we're talking about the model high -- you're saying the model has to be stuff you know in advance. What kind of

R-square can you expect, and I can tell you what you can expect. You can just pick real low R-squares here, on the order of, are you ready? Point-two or lower, .1, .2, I'd be surprised.

There, what you're talking about, you're talking about age and gender and stuff like that. You're not talking about the big drivers and things that actually cost, which is hospitalizations.

MS. CLARK: The HCC ones, I've gotten some that have been the highest, around .3, I think.

DR. WEINTRAUB: Yes, so, there you go, the other one is .3. You're predicting that 30 percent of the variability in costs.

Now, you know, Jeptha's heard me go through this kind of stuff before, do you believe a model like that, and one of the responses that Ronald Crumhold got to this, well, it's gives lots of -- well, that means you have lots of room for variability of your

providers, and you can say, that doesn't matter at all, you know, if you can't predict costs, then just use average costs, and it doesn't matter.

But if that's the case, then you really -- then I don't believe that, at all, I mean, what you'd like to see is that providers help determine that, and I guess one of the things you could do, in looking at this, is looking -- if you add in providers to the cost, does that add to variability in a validation sample?

I mean, there are things you can do to try and get at this, but I think this is -- that the ability to truly risk adjust here is going to be pretty minimal.

CO-CHAIR CURTIS: I think that's why we didn't see the results in any of the applications.

DR. WEINTRAUB: Maybe so. I don't want to know. But should, not just in terms of this kind of modeling, but also, when you're

using modeling where there are discriminations with the C-index.

Should NQF be setting some kind of standards? I realize that goes beyond this panel, but carrying it forward is something to think about.

MS. TURBYVILLE: I think right now, the most recent guidance is in that testing task force report, that came out at the end of last year.

So, but you're right, that's something for us to think about, and take back, as we continue to build our guidance for the expert panel.

So, I did want to make sure that we open up for public comment, before most of you dart out of the room, just in case it's something they would like feedback from all of you.

So, Operator, please, at this time, could you open the lines for public input or comment?

OPERATOR: Certainly, that is *1 for public input or comment.

We have no one in queue at this time.

MS. TURBYVILLE: So, Jeptha, it's three o'clock, now. So, we're suppose to end at 3:30 p.m. today. Should we wrap up with next steps?

CO-CHAIR CURTIS: Yes, so, I think we should defer in the --

MS. TURBYVILLE: Or did you want to go into --

CO-CHAIR CURTIS: -- process going forward, we have seven more measures. We've gotten through --

CO-CHAIR ROSENZWEIG: It's seven more.

CO-CHAIR CURTIS: -- seven, in two days, which is sobering, and probably, is useful for you guys to reflect on further ones.

But I don't know what the worst

case was, but that's pretty close to my worse case.

MS. TURBYVILLE: Worst case scenario in this situation was zero.

CO-CHAIR ROSENZWEIG: So, we did seven, is that right?

MS. TURBYVILLE: That's right, congratulations.

CO-CHAIR ROSENZWEIG: And there were 14 on the list?

MS. TURBYVILLE: Right, and so, we've hit every single vendor within this group. So, hopefully, as we did with the ABMS measures, we'll continue.

CO-CHAIR CURTIS: So, I think there are two parts that I think we should cover.

First, next steps, how we're going to get through the additional measures, and then secondarily, kind of just pause for a reflection from the members here, the TAP members, as to is there any way that we could

refine this process, as we do it?

I mean, we're kind of, I think, stuck with what we're at, in terms of the criteria, for assessment, but process-wise.

MS. WILBON: So, operationally, we've got, as Jeptha said, we've got seven measures left.

One is an NCQA measure, which is for relative resource use of people with cardiovascular conditions.

We've got one, two, three ABMS measures left, two on CHF and one on -- I'm sorry, two on CHF and one on CAD, and then we've got another three Ingenix measures.

So, I guess it depends on -- we've got a couple of ways to address it. We're definitely going to need some follow up conference calls, so, what we'll be doing, if not by the end of this week, by early next week, sending out an availability survey to you guys, to probably schedule, I'm going to start with three conference calls over the

next month, to try to get through as much as possible, so, we can start filtering -- well, we can probably start filtering some of this, probably just right now, the NCQA measures.

The only one that we didn't really ask for a lot of follow up -- we have to check our notes, but to see what we can start filtering to the Steering Committee, for them to get through, and I could -- we can kind of ask the Co-Chairs of the committee, how they would like to kind of chunk those out.

Do you want to start with the NCQA measure, or start with the ABMS measure, since there seems to be a little bit of kind of comfort with those now, and then save the Ingenix measures for last, or how do you guys want to try to address those?

CO-CHAIR ROSENZWEIG: Can I just ask? What is the deadline, in terms of presenting of this material to the Steering Committee?

MS. WILBON: Right, so, the

Steering Committee meets at the end of June, and that meeting at June 29th and 30th, that meeting is a two day meeting and our goal for that meeting was to have them review everything from this meeting, from this group.

So, the focus of that meeting is only this -- just this TAPs work.

CO-CHAIR ROSENZWEIG: Just this task force?

MS. WILBON: Yes.

CO-CHAIR ROSENZWEIG: Well, that is going to be difficult. I mean, if you think about seven measures and if we do this over, let's say, if we schedule conferences calls to do them, I mean, you can't really expect a conference call to last all day.

MS. WILBON: No, absolutely, but we do two hour --

CO-CHAIR ROSENZWEIG: So, the maximum for conference call would be two hours.

MS. WILBON: Yes.

CO-CHAIR ROSENZWEIG: And that might be two measures.

CO-CHAIR CURTIS: So, we have a month? A month to do this, and two hours is ambitious, although if we're going to do it, we're going to do the ABMS measures as a group --

MS. WILBON: First, okay.

CO-CHAIR CURTIS: -- and we're going to do the Ingenix as a group, you know.

MS. WILBON: Okay.

CO-CHAIR ROSENZWEIG: Can we suggest, that we at least finish CAD and ask that the Steering Committee delay its consideration of the other clinical conditions, like stroke and CHF?

MS. WILBON: Well, what --

CO-CHAIR ROSENZWEIG: To a later meeting?

MS. WILBON: What we would do, I mean, ultimately --

CO-CHAIR ROSENZWEIG: I'm getting

a dirty look from her.

MS. WILBON: If we can't get through, we can't get through. We would send the Steering Committee as much as we could, by the time -- that meeting is already scheduled. It's in the works.

We can't really delay that work, but we would give them as much as we can. We're going to try to give them about a month or so, at least three weeks, to review what you guys have done, and I suspect that even with -- even if we give them four or five measures, that it may take them as long, if not longer, to get through them.

So, even if we had, honestly, if we gave them all 14 measures, I'm not sure that they would get through them all in a two day meeting.

So, we can talk. We'll probably need to talk a little bit more internally with the team, to figure out what is the strategy for that.

But I think it is reasonable if we could give them at least half of the measures by -- to review at the June meeting.

CO-CHAIR CURTIS: Well, we've already done that.

CO-CHAIR ROSENZWEIG: We've already done that.

MS. WILBON: Right, but there is still some follow up and you know, some potentially re-voting. So, that takes time, as well.

CO-CHAIR CURTIS: Do you think the measure developer -- I mean, just based on prior experience, have measure developers been able to respond and have follow up TAPs within a month?

MS. WILBON: Yes.

CO-CHAIR CURTIS: My recollection is that it's usually a little bit -- I know there are time --

MS. WILBON: It depends. I think the type of -- a lot of the information we're

asking, should be relatively -- shouldn't take them that long to respond with.

So, they should already have these R-squares and they shouldn't -- the things we're asking, not to re-test or, you know, it should be clarification. Most of them are clarifications, or things that shouldn't require weeks to --

CO-CHAIR CURTIS: So, I think people are just starting to realize that this is a full-time, but unpaid job.

DR. WEINTRAUB: Most of us have three or four of those already.

CO-CHAIR CURTIS: Yes.

CO-CHAIR ROSENZWEIG: Yes.

CO-CHAIR CURTIS: Well, I think what we should do is sort of set, what is the expectation for participation before June 10th, and I think it would be reasonable to, you know, not reasonable, but the highest that I would feel comfortable committing to is like three, two hour conference calls, and I think

beyond that, you're really pushing the boundaries of both good will.

MS. WILBON: I think that is reasonable.

CO-CHAIR CURTIS: I don't know if others --

DR. HWONG: I just want to point out that there is Memorial Day weekend, kind of at the end of May, just to be cognizant of travel plans.

CO-CHAIR CURTIS: And also, like, what is the quorum that's going to be like, getting nine or ten people together, for two hours, three times in the next three weeks?

DR. WEINTRAUB: It's not going to happen.

CO-CHAIR CURTIS: It is going to be difficult.

CO-CHAIR ROSENZWEIG: It's going to be impossible, yes. I think we could probably get one follow up conference call, but three by the end of June?

CO-CHAIR CURTIS: I don't know.

MS. TURBYVILLE: Before the end of June.

CO-CHAIR ROSENZWEIG: Before the end of June?

DR. WEINTRAUB: It can't be done.

MS. BOSSLEY: I think we need to just let us spend a little time thinking through, because we have a better sense of what we'll take and what you need, to run through these measures.

So, give us a little time to huddle and we'll come up with a plan for you.

MS. WILBON: A reasonable plan.

MS. BOSSLEY: Yes, sure.

CO-CHAIR ROSENZWEIG: So, could you just run, mention the ones that are still left to be done?

MS. WILBON: Yes, we have the NCQA, RRU, for cardiovascular conditions, relative resource use for people with cardiovascular conditions.

CO-CHAIR ROSENZWEIG: Okay.

MS. WILBON: Fifteen-seventy-two, which is the episode of care for management of --

CO-CHAIR ROSENZWEIG: CAD?

MS. WILBON: -- coronary artery disease, which is from ABMS, 1574, which is episode of care for CHF over 12 month period, from ABMS, 1575, episode of care for management of post-hospitalization CHF over a four month period, from ABMS, ETG based CHF, from Ingenix, 1591, 1594 is ETG for CAD, from Ingenix, and 1596, ETG stroke from Ingenix.

MS. PARKER: And my recollection on the last one, the 1596 was that there was going to be some discussion among the lead discussant, as well as maybe the rest of the group, that was to review that, based on its applicability, to the same criteria as the AMI.

CO-CHAIR CURTIS: Right, so, we'll follow up --

MS. PARKER: Is that correct?

CO-CHAIR CURTIS: -- with the measure developer and the NQF staff about that.

MS. PARKER: Okay.

MS. WILBON: Okay.

MS. TURBYVILLE: You still need someone on the TAP's input on it, so, yes --

MS. PARKER: That will be the discussant, correct?

CO-CHAIR CURTIS: Right.

MS. TURBYVILLE: We'll see if we can take it offline and see what is going on with that.

DR. HWONG: And the only one thing I'd mention, in terms of like, yes, time frame and whatever, but you know, having -- I'm sort of the lead reviewer on the CHF version of the Ingenix, you know, Ingenix CHF ETG and because it is that same kind of episode, excuse me, the one year episode concept, a lot of it, at least when I was looking at it, it looks extremely similar.

So, I mean, you know, hopefully, maybe if we can emphasize like, when we get on these calls, just time saving like, really, you know, just, even if we had like, sort of the write-ups or something from like the previous voting, just to kind of have the lead person go through and say, "Yes, that is the same, that is the same," you know, here is where it might be a little interesting or different, if at all, you know, and then --

CO-CHAIR CURTIS: I think Brenda did a nice job with that approach for the --

CO-CHAIR ROSENZWEIG: Yes.

MS. PARKER: Thank you.

CO-CHAIR ROSENZWEIG: Yes, it's really up to NQF to decide what order they want us to do these, but I would suggest that we try to get the CAD one completed, you know, at least --

CO-CHAIR CURTIS: Do we have the diabetes ones completed?

CO-CHAIR ROSENZWEIG: The

diabetes, yes, get the diabetes and CAD and MI ones completed, and because once we get into CHF and we get into stroke, we're dealing with new disorders, so, probably a lot of additional things.

So, I would --

MS. TURBYVILLE: However, that might --

CO-CHAIR ROSENZWEIG: Just consider that.

MS. TURBYVILLE: Yes, we'll definitely consider that, but it could break up with the vendor approach on the Steering -- on the conference call.

But we'll play around with it. We'll bounce it off of you guys. We'll come up with a strategy and in the very near future, so that we can bounce if off of you guys, as we prepare for these calls.

MS. BOSSLEY: Yes, it will depend on whether the developers are available too, and there is no point in having a call to

discuss the measures, if they're not there.

So, we have to factor all of that in and --

CO-CHAIR CURTIS: It might be fast.

MS. BOSSLEY: It may be fast, but then you have a lot more comments to deal with on the back end. So, one way or the other, you're going to have to deal with it.

CO-CHAIR ROSENZWEIG: Do we come from lots of different time zones? Are we all from the east?

MS. BOSSLEY: David, you're in California?

DR. PALESTRANT: Hello.

MS. BOSSLEY: L.A.?

CO-CHAIR ROSENZWEIG: Well, that creates problems, then. I mean, that means an evening call.

DR. PALESTRANT: There are certain times that I can do it late morning, or at least, I can work with you guys.

CO-CHAIR ROSENZWEIG: I've been on many 7:00 to 9:00 p.m. calls.

CO-CHAIR CURTIS: Anyway, I just want to thank the members of the TAP and the NQF for doing a wonderful job of getting us as far as we've come, and obviously, as I predicted, it's been intense, and continuing, ongoing.

MS. WILBON: Thanks to our Co-Chairs, too, for helping us plow through and get through as much as we did. I know Jeptha was a little scared, unsure about how this was going to go, but I think we actually did a really good job.

This is brand new, as we said, so, great job for plowing the way.

(Whereupon, the above-entitled matter concluded at 3:06 p.m.)

G

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Technical Advisory Panel

Before: NQF

Date: 05-11-11

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R. Gross

Court Reporter