NATIONAL QUALITY FORUM

+ + + + +

RESOURCE USE STEERING COMMITTEE

+ + + + +

TUESDAY
AUGUST 30, 2011

+ + + + +

The Steering Committee met, at the Venable LLP Conference Center, 575 7th Street, N.W., Washington, D.C., at 9:00 a.m., Bruce Steinwald and Tom Rosenthal, Co-Chairs, presiding.

PRESENT:

THOMAS ROSENTHAL, MD, Co-chair
BRUCE STEINWALD, MBA, Co-chair
PAUL BARNETT, PhD, VA Palo Alto Health Care System
JACK BOWHAN, Wisconsin Collaborative for Healthcare Quality
JEPTHA CURTIS, MD, FACC, Yale University School of Medicine*
KURTIS ELWARD, MD, MPH, FAAFP, Family Medicine of Albemarle
LISA GRABERT, MPH, American Hospital Association
ETHAN HALM, MD, MPH, University of Texas Southwestern Medical Center*
THOMAS LEE, MD, Partners HealthCare System, Inc.
JACK NEEDLEMAN, PhD, FAAN, University of California, Los Angeles School of Public Health
DORIS PETER, PhD, Consumers Union*
STEVE PHILLIPS, MPA, Ortho-McNeill-Janssen Pharmaceutical, Inc.
DAVID REDFEARN, PhD, WellPoint

BARBARA RUDOLPH, PhD, MSSW, The Leapfrog Group
JOSEPH STEPHANSKY, PhD, Michigan Health and Hospital Association
JAMES WEINSTEIN, DO, MSc, The Dartmouth Institute for Health Policy; Dartmouth-Hitch Clinic*
DOLORES YANAGIHARA, MPH, Integrated Healthcare Association

NQF STAFF:

CARLOS ALZOLA, Consultant*
TAROON AMIN
HELEN BURSTIN, MD, MPH
LAURALEI DORIAN
SARAH FANTA
SALLY TURBYVILLE, MS, Consultant
ASHLIE WILBON

ALSO PRESENT:

DAN DUNN, PhD, Ingenix
BEN HAMLIN, MPH, NCQA
TOM LYNN, Ingenix
JANET MAURER, MD, MBA
KIMBERLY RITTEN, HealthPartners
PATRICIA SINNOTT, PT, PhD, MPH, VA Health
CHERI ZIELINSKI, Ingenix

* Participating by teleconference

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

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

9:10 a.m.

MS. WILBON: So, good morning, everyone. Thank you for coming. We=re glad to see everyone back again and that we didn=t scare everyone away from the last meeting. And again, it=s summertime, so glad that you guys were able to make it down.

This is the second and final Steering Committee meeting for this project. So, we are going to be looking for some of your insights on day two to kind of wrap things up and input on how we can move forward to the next steps.

For this morning, we are going to start again with brief kind of introductory slides to get everyone started.

We will start with a brief introduction of everyone for the record, so that we have an idea of who is in the room and who is on the phone.

A couple of other housekeeping items. I think everyone has been in this building before, but restrooms are outside to the front lobby area and then over to the right.

Everyone, if you have a laptop, we do have thumb drives with electronic versions of all the documents we have sent. I think a lot of you have them from email, but we also have it on a thumb drive, if you need it. You have a folder of documents we will referring to throughout the two days.

And I think that=s it. So, let=s start with some introductions from around the room and on the phone. Let=s start with Steve at the end of the table.

MR. PHILLIPS: Yes. Hi. Steve Phillips with Johnson & Johnson, and I don=t have any conflicts to declare.

MS. WILBON: Actually, we don=t have to do conflicts this time.

MR. PHILLIPS: Oh, okay.

MS. WILBON: You=re ahead of the game, but we don=t need it this time. So, great.

MR. PHILLIPS: Okay.

MS. WILBON: We=ve done it so many times at this point; we just didn=t need to do it again.

DR. BARNETT: I=m Paul Barnett. I=m with the Health Economics Resource Center in the Department of Veterans Affairs.

DR. STEPHANSKY: Joe Stephansky. I=m with the Michigan Health and Hospital Association.

DR. RUDOLPH: Barb Rudolph. I represent the Leapfrog Group and the National Association of Health Data Organizations.

MR. BOWHAN: I=m Jack Bowhan, Wisconsin Collaborative for Healthcare Quality.

DR. REDFEARN: David Redfearn, WellPoint.

DR. BURSTIN: Helen Burstin, NQF.

CO-CHAIR STEINWALD: Bruce Steinwald. I=m an independent consultant and not so recently anymore with the Government Accountability Office.

CO-CHAIR ROSENTHAL: Tom Rosenthal. I=m the Chief Medical Officer at UCLA in Los Angeles.

MS. WILBON: Ashlie Wilbon, Senior Project Manager for NQF.

MR. AMIN: Taroon Amin, Senior Director, NQF.

MS. TURBYVILLE: I=m Sally Turbyville with Impact International.

MS. FANTA: Hi. Sarah Fanta, Project Analyst with NQF.

MS. DORIAN: Lauralei Dorian, Project Manager, NQF.

MS. YANAGIHARA: Hello. I=m Dolores Yanagihara with the Integrated Healthcare Association in California.

MS. GRABERT: Lisa Grabert, American Hospital Association.

DR. LEE: Thomas Lee from Partners Healthcare and Harvard Medical School.

DR. NEEDLEMAN: Jack Needleman from the UCLA School of Public Health and the UCLA Patient Safety Institute.

MS. WILBON: Thank you.

Helen, did you want to give a brief introduction?

DR. BURSTIN: I want to add my welcome. Sorry, she asked me if I wanted to say hello just as my mouth was full of some very yummy yogurt and granola, which I don=t think is actually very low fat or low sugar, as I tasted it.

(Laughter.)

If children would like it, that=s not a good sign. It kind of tastes like Trix kind of cereal yogurt.

Anyway, welcome.

Are there any other Members? Okay.

So, anyway, just welcome.

For those of you who have seen it, and I didn=t see it posted yet this morning, the first phase report should be posted today. I read it last week. I thought it was just a phenomenal piece of work. I thought the team did a great job. You guys did a great job. It just really summarized the issues. It crystallized it so well.

Some of you probably saw that demos were announced last week for CMMI, the various payment demos. And there was this question that arose about, well, is NQF on-track to really help with some of the payment measures that they are going to need there? And it was sort of an interesting issue.

I think we are starting down that path, but at the same time it is very obvious from reading that report that we really need those demos to actually help us understand what the standardized measures should be. So, I think, as we always thought, this is a great opportunity for learning. Hopefully, it is

sort of the first phase of this work, but by no means the end.

So, thank you for all your insights. I thought it was just a great piece of work, really well-written. I think we have learned a lot with your help. Thank you.

MS. WILBON: And we will go back to introductions of people on the phone. Sorry, I skipped over you.

Tom, if we could have the people, I guess, on the speakers= line and then move over to the participants= line for people to introduce themselves?

THE OPERATOR: All lines are open. We do have Cheri Zielinski.

MS. ZIELINSKI: Hi. Cheri Zielinski with Ingenix. I=m happy to be here. Thanks for having us.

THE OPERATOR: Another speaker is Jeptha Curtis.

DR. CURTIS: Hi. Jeptha Curtis from Yale.

THE OPERATOR: And another speaker, Tom Lynn.

MR. LYNN: Yes, Tom Lynn from Ingenix.

THE OPERATOR: And we do have a participant from HealthPartners.

MS. RITTIN: Yes. Hi. This is Kim Ritten from HealthPartners.

THE OPERATOR: And that is your on-the-phone audience.

MS. WILBON: Okay. Thank you.

So, we are going to actually just jump right into the slides for today. We are going to do a very just kind of brief introduction.

Today we are going to start out with a discussion on the Ingenix measures. There were some changes in their specifications through this last kind of developer measure update phase that we do after each of the meetings. So, we will brief everyone on that and then have a discussion on

that.

And then, we will move right into the evaluation and final recommendations for the seven remaining measures that are from the Pulmonary and Bone Joint TAPs.

And then, day two will be -- hopefully, by today, we will finish all the measure review stuff, and then, by day two, we will move into some of the more kind of reflection on evaluating the measures and kind of next steps on how we might move forward with some future efforts.

So, that is our agenda for the next two days.

Just a quick project update. As Helen already said, we did post the draft report for the Cycle 1 measures for public and member comment, starting today, and that goes through September 28th. Included in that report are these four measures: the two HealthPartners measures, the total resource use and total cost of care, and then the two

measures from NCQA for diabetes and cardiovascular conditions.

So, because of this discussion for the Ingenix measures, we are holding those measures off for the Cycle 2 report until we can kind of make sure we have resolved all of those issues. And as you know, all the TAP meetings have been complete.

I won=t spend time on this, but you have this packet of slides in your folders, if you want to kind of look at where we are with the timelines for both cycles of measure review.

And we are going to just kind of jump right in this morning and talk about the kind of latest development with the Ingenix measures.

So, as I mentioned, as we were going through this process, obviously, it was a learning process for us all. If you recall, when you evaluated the HealthPartners measure way back in the beginning, I think we started

that conversation on the phone some time ago, and there were two costing approaches that were proposed in the measure for both standardized cost and for actual prices. The Committee felt at that time that those two costing approaches should be split out of the measure, and two separate, individual measures should be submitted from HealthPartners. So, they did that. You evaluated both measures independently, one for actual prices and then one for standardized prices.

What we didn=t realize at the time is that Ingenix had a very similar approach in their measure, but we didn=t actually catch onto that fact until much later in the process, which was, I think with all the measure review and all the meetings that we had, it took us a while to kind of catch onto that, until we actually had this kind of quality check, and going back through our process to make sure we were being consistent.

What we did is we went back to

Ingenix, brought it to their attention, and gave them the same option that we did for HealthPartners, which was to either pick one methodology that you apply to all your measures for either standardized pricing or actual cost or split each of their measures into two measures.

And what they decided to do was to apply actual cost to all their measures. What that means is, for the four measures that you already voted on, those measures were voted on having both costing approaches in the measure.

So, we just kind of wanted (a) to bring that to your attention. Two, since they have now applied the single approach, to determine from everyone here whether or not you believe that that fundamentally or inherently changes the measure and whether or not you think that requires more discussion or how you would like to move forward on that.

So, I will kind of let Tom and Bruce take that over.

CO-CHAIR ROSENTHAL: So, Ashlie, what are our options in relationship to the question? So, the problem, to the extent that there is one, is that Ingenix did not specify the two costing methodologies, and we voted on all four of them with that being the case?

MS. WILBON: Right, with an Aeither/or@.

CO-CHAIR ROSENTHAL: With an Aeither/or@. Whereas, NQF had insisted that the other submitters clarify. So, what are our options as a Committee in relationship to this?

MS. WILBON: So, the options would be, if you guys feel that them changing their costing approach to actual prices only does not change the measure, and that your votes would still be the same on that measure, now knowing that it only has one costing approach, a single costing approach, then that=s it. We would just say the Committee does not feel that that intrinsically changes the measure,

and the votes would stand.

If you do feel that it changes the measure, and that your vote on the measure or your kind of ratings on the measure might change, then we would have you guys revote on the measure today, on those four measures.

CO-CHAIR ROSENTHAL: So, does everybody understand the problem and the options available to us as a group? Because we should at least clarify the question before we discuss it. Everybody get the question?

Barbara?

DR. RUDOLPH: So, will our prior vote then be eliminated if we choose to?

MS. WILBON: Yes.

DR. RUDOLPH: And with that, then, for the end-users, if they use standardized pricing, they would not be following the NQF endorsement?

MS. WILBON: Right.

DR. RUDOLPH: Is that correct?

MS. WILBON: Right.

DR. RUDOLPH: Okay.

CO-CHAIR ROSENTHAL: So, they switched from giving the option of a standardized price or the actual to submitting them with only the actual prices? So, if it was a total PMPM -- or what were the four measures again? Let=s again be sure that we=re --

MS. WILBON: I know the slide is a little bit smaller because we had two screens.

CO-CHAIR ROSENTHAL: It=s hard to see from out there.

MS. WILBON: But it is the 1591, which is the ETG-based, non-condition-specific, and they are all now cost-of-care measures. And 1591 was an ETG-based CHF, cost-of-care; 1595 is ETG-based diabetes, and, then, 1594 is ETG-based coronary artery disease.

CO-CHAIR ROSENTHAL: So, if we took coronary artery disease, for example, then, with actual pricing, if it was, say,

Massachusetts versus Minnesota, the cost of care would be $8,000 per person in Massachusetts and $7,000 per person in Minnesota versus some standardized pricing that said it was really a bread basket of utilization. So, it is utilization versus dollars.

Tom?

DR. LEE: So, my assumption is that they will have no way of dealing with risk-sharing-type contracts, which are becoming very common in Massachusetts, 50/50 risk-sharing, that it will simply overwhelm the methodology and be not very useful.

I mean, if that is true, then I think that is a problem because I think that at the end of the day we want to help people understand when they are utilizing more than other folks, even if they are in a risk-sharing arrangement.

DR. REDFEARN: I have to confess that I must have missed something because I

can=t change my vote because my assumption was they were using real prices all the way along.

And I recall we had a lot of discussion about the issue of are real prices comparable across geographies. We discussed that at length in part of our evaluation. I don=t remember the synthetic pricing at all. I=m sorry, I must have missed something --

MS. WILBON: That was in relation to the HealthPartners measure. So, the HealthPartners total cost measure is the one that we discussed at the last in-person meeting, but it wasn=t brought up in the context of the Ingenix measure. So, that is one of the reasons why we wanted to bring it to your attention, because of that in the context of that discussion that you had.

CO-CHAIR ROSENTHAL: Yes, I have to confess, David, my recollection of the Ingenix measures was I thought they were standardized prices.

(Laughter.)

Hence, the confusion and, hence, the question.

Because, frankly, if we were going to be completely consistent as a group, then we would say, well, we voted to accept these with either costing methodology, one could argue that, well, we assumed this one was okay, so why would we need to revote? And in my head, the only reason potentially to revote was the notion that there might have been some confusion. And in my head, I thought all the Ingenix ones, in fact, had standardized pricing and did not have the option of dollar pricing.

And in relationship to the HealthPartners one where we had a pretty extensive debate about whether dollars were okay, making comparison to the efficiency opened a question because, well, the question, as Tom it, is it really efficient if, in fact, you haven=t accounted for wages, for example? And how would you, then, fairly compare

Minnesota with Massachusetts, or whatever?

Jack?

DR. NEEDLEMAN: Yes, the basic thrust of the conversation we had about standardized pricing versus actual prices or actual payments, because it wasn=t what was charged, it was what was paid that went into the estimate of resources, is that each measure provided some information of value. And depending upon what your use was going to be, standardized might be more useful than actual revenues received for services, while in other cases the actual revenues received would be a more useful measure.

So, I actually like the option of having both, and not necessarily in the measure definition. But if I were an Ingenix client, I would be wanting to receive my data both ways.

And so, one of my questions is whether Ingenix plans to continue to offer the standardized pricing if that is not the

endorsed measure basis for estimating resources.

CO-CHAIR STEINWALD: Maybe since we have two Ingenix people, can you respond to that question?

MR. LYNN: Sure.

DR. DUNN: Hey, Tom, I=m here as well.

MR. LYNN: Oh, sorry, Dan.

DR. DUNN: Why don=t I take it? This is Dan.

MR. LYNN: Yes, please.

DR. DUNN: Yes, Jack=s point I think was right on. Our customers in many ways would like to see it both ways. One is where standard pricing is enforced. I think as someone noted, it has become a weighted utilization approach. It removes differences between hospital contracts or fee schedules, different parts of the state.

And then, to the point of the actual amounts, which may reflect in some

cases decisions on which hospitals to use or which center to go to for an MRI, or whatever. That is part of decision. It is part of the dollars spent on healthcare.

We have offered an option. Again our measures are used even outside of the real timing afforded the customer using either approach. I think there is value in both. To be honest, in the majority of the cases where these measures are using physician measurement, it is usually within a market or a state even with actual prices.

Standard pricing is much more in some ways the atypical case, at least in practice right now. It is usually used where you have something like Wisconsin where there is a data aggregation.

And one of the reasons they removed the real prices is because of the confidentiality, same thing with NCQA and RRU measures. One of the main drivers of moving to standard prices was, in fact, the fee

schedule of the data submitted, and then, also, obviously, the way they were doing comparisons across those states, across health plans. So, there was a need to equalize pricing.

CO-CHAIR ROSENTHAL: So, could I pose the question to Ingenix maybe slightly differently? Is it the proprietary nature of the prices that cause you to put the measure forward as a dollar-only proposal as opposed to a standardized pricing proposal? Or is there some other reason why you selected the one that you selected?

DR. DUNN: To be honest, it is actually was, given the amount of time involved, it would have taken us to put all the standard pricing logic in tables into a format that was acceptable for NQF. That was the main driver there. In the future, if we have the opportunity and time, we would be happy to submit the standardized pricing approach.

CO-CHAIR STEINWALD: ***So, it was basically to simplify your life and submit one measure that you felt reasonably good about, and it is very separate from your business model that enables clients to select which kind of pricing methodology, or both, to suit their own needs? So, it was a really separate decision of what to submit to NQF, based on a different set of criteria than what you are offering to your clients?***

DR. DUNN: ***Right. So, maybe to summarize, our preference is to provide flexibility to the customers because in many ways the standard pricing is an important part of the measure, to use this standard pricing. But the clinical methodology is the same whether you use actual prices or for standard prices. So, that flexibility I think was always in our minds.***

But given the change in preference of NQF, if we are going to support standard pricing in the measure, to include that part

of the methodology. And again, given timing and our ability to pull that together quickly, that is why we chose to submit based on the actual prices, actual cost.

So, did I answer the question?

DR. STEPHANSKY: Well, I am more puzzled than ever in terms of us endorsing a set of measures or a measure set where it is not the way that it would be used in practice. I am wondering, do we really need to separate these out like we did for HealthPartners? Or can we go ahead as we started before?

CO-CHAIR ROSENTHAL: Well, I think they are submitting it now as prices only. So, we will have to vote to either -- I think we will have to sort of make a judgment as to whether or not -- we can pose the question one of two ways. We are going to revote de novo on the overall acceptability of the measure or we can vote to affirm our prior decision that, if it had both methodologies, that we are reaffirming our prior decision to accept it.

I think, either way, we are going to be determining whether or not we believe the measure as submitted now is acceptable as an endorsed measure. And it is only as it is, and it doesn=t matter what they offer their customers. It=s interesting, but it doesn=t matter what they offer their customers.

MS. WILBON: So, Joe, to kind of piggyback on your question, even if we did decide to go back to Aeither/or@, they would still have to specify their standardized -- even in the original submission, if you recall back to like the NCQA measures and the HealthPartners, they actually specified what their standardized pricing approach was. They gave access to standardized pricing tables. So, that work would still be required on behalf of Ingenix to specify that in the measure.

CO-CHAIR ROSENTHAL: Is there further discussion? I don=t want to belabor the point, but perhaps, Jack, you could give

the 25-word or the two-minute version of why the prices are a good measure nationally, because, again, this is a national measure, not a local measure or a regional measure. This is a national measure.

And, Tom, you could perhaps give the two-minute version of your concerns about not factoring prices and the potential problems of holding providers accountable for factors over which they have absolutely no control.

So, maybe we could just do that for two minutes, and then we could call the question.

DR. NEEDLEMAN: So, part of the reason, I think there are two fundamental reasons why a price-based measure is a reasonable one and you would want to see that data.

And No. 1 is that is the way everybody else reports. That is the way the data gets routinely reported. So, for most of

the other measures of cost we have, the regional variations in cost are reported. And one has to, then, back that up to think about what the differences in utilization are versus prices across regions. That is important information. So, having the price-based measures rather than standardized price measures tells you something.

The other reason is, as an economist, if there are major differences in relative prices of services, we should expect to see differences in the mix of services. Let me think. If surgery for a specific procedure for some reason is much less expensive in Arizona compared to physical therapy and non-surgical interventions than it is in Virginia, we would expect to see more surgery in Arizona than Virginia, and we would explain that in part by the difference in prices, the relative cost of taking path A rather than path B.

So, the price information has

important information to understand incentives in the system and why people are making decision to pursue different kinds of treatment, if the relative prices of different kinds of treatment vary from state to state or region to region.

DR. LEE: And I do think price matters and Jack=s point is well-taken. I also know that my perspective is distorted by being in Massachusetts, which is in a different stage of development in healthcare from the rest of the country.

I do fear that measures just based on price will tell the world that real estate is more expensive in Massachusetts than it is in North Dakota. I think the world already knows that, and I am not sure that measures that primarily convey that information are going to be that helpful. You know, real estate translates into higher wages, and so on.

In the world in which I work,

frankly, it is getting to the tipping point now that most commercial business is in risk-sharing, 50/50 risk-sharing. I think it is the right direction. Prices based upon fragmented units of service, I hope will become less relevant to the country as a whole and it will be more about what happens to populations over time.

I think that, as you try to improve your efficiency with populations, what you are really interested in is who is doing better than you in the number of units of service that patients with certain conditions are getting. So, for the learning perspective of providers, the standardized price approach is more valuable.

CO-CHAIR ROSENTHAL: Yes, I would have to add just my two cents on this because, Jack, I think the points are well-made about the value of the raw numbers. But in the geographic variation discussion that occurred over the last two years around Medicare, it,

frankly, led to some very distorted conclusions. And that is the biggest concern I=ve got.

It was quite clear that Congress was ready to act, from my observation, to do things without factoring in the prices, basically, calling the providers in particular regions inefficient, and this is supposed to be an efficiency measurement, that had nothing whatsoever to do with provider efficiency or inefficiency. It had only to do with prices.

And so, the potential misuse of the price-only data is the concern I have. If people were all wise and thoughtful in the way that you are, I would agree completely about the value of putting price, dollar-denominated figures, out there that say the hospitals in Massachusetts cost more than the hospitals in South Dakota. And then, use the very thoughtful analytics that you would apply to saying here=s why and we understand why, and it doesn=t mean anything or here=s what it

means.

But the fact is, the way it has been used for policymaking up until now has been those providers in Massachusetts are grossly inefficient and somehow they should be punished or others rewarded because of the, quote, Ainefficiencies@ that, again, have nothing to do with the actual provision of services. So, it is the misuse, potential misuse, that troubles me about this.

CO-CHAIR STEINWALD: Hang on a second.

(Laughter.)

Medicare is my beat. And Medicare routinely uses standardization in almost everything they do.

But a big issue for Medicare, and it is the topic of an IOM committee that I participate in, is how they do the standardization. And that is one of the advantages of using actuals, is you know what they are. They are what is actually paid.

Once you get into standardization, theoretically, it makes across-geographic-areas comparisons more valid, but you run into all sorts of technical issues about how to do the standardization.

CO-CHAIR ROSENTHAL: Well, I get it, but the IOM, as I understood it, in their very first run-through -- and again, there may be debates as to whether the wage adjuster for comparing State A to State B was accurate -- there was, I thought, widespread agreement that the original raw scores of showing the amount of variation, it scrunched up rather significantly.

And if what you are trying to compare are the provider efficiencies of providers in one place versus another, once at least a run at standardizing the prices was done, the amount of variation was considered much more believable than it was with the raw scores. The raw scores were viewed as, well, this isn=t valid because they haven=t made any

attempt to take the prices into consideration.

DR. NEEDLEMAN: Yes, I think there is information in both measures. The back-and-forth that we have been having just is an echo of the earlier conversation we had about the value of each and what one could learn from looking at each.

I am actually deeply disappointed that Ingenix did not come back with paired measures and say, AWe would like the pair endorsed and we expect to use them as a pair. We expect to sell them as a pair because there is value in each.@

And we saw they have a standardized pricing methodology. We know that because they have used it. They chose not to do the work to create a separate application with it. Others that have submitted measures to us have.

So, I am deeply disappointed in the way this was approached and what we have got here. So, the issue is, do we wait until

they come back and say, AGive us two.@? Do we say, AGive us two paired and we=ll look at it then@ before we approve? That=s an option.

DR. BARNETT: So, the issue before us is whether we revote these four, right? So, my recollection is that we turned the ball down, is that correct?

MS. WILBON: We have the results, but we were kind of holding off on sharing. We decided we didn=t want to kind of taint the --

CO-CHAIR ROSENTHAL: But the results are relevant to this discussion.

MS. WILBON: Yes. Okay.

DR. BARNETT: So, I=m not sure if, either or both, that I would still be in favor of any of these measures. And I think it is just that there is not a very good fit with how we set up this to either have a non-condition-specific or a condition-specific measure. And the Ingenix is something a little bit different.

So, the problem with the Ingenix non-condition-specific measure, as I see it, is that you have this very complicated episode-grouping software which is actually not needed to a non-condition-specific measure. It=s too complicated. So, even if they had one or the other, you really wouldn=t go to all that trouble to do episode groups to come up with this non-specific measure of efficiency.

And the problem with the other ones is, similarly, you have to create episodes for everything in order to come up with a CHF measure or a diabetes measure. And so, again, it is more complicated than is needed.

So, regardless, I don=t think the costing issue is really what determines the decision. It is just that it is not a very good fit for what NQF is trying to get out of this process. So, that is my take on it.

CO-CHAIR ROSENTHAL: Well, I

personally think that the issue of what was the vote seems to be somewhat relevant to this, does it not? That=s what you are saying.

So, why don=t we put what we voted. And then, we can decide one by one, we can decide in aggregate. We can do this any way the group decides they want to do it.

It is kind of small. All right, I think I can read it here.

Let=s do them one at a time. So, we will do the non-condition-specific one first. That=s 15 -- I can=t read it; it is the top one on there -- 1599. Thank you.

The overall recommendation was 12 yes and 6 no. The feasibility vote was --

MS. WILBON: Oh, right, if you recall, with that measure we had split up the discussion on that because we have the pricing tables as a separate discussion, I think it was on a call. So, after that call, we had you guys vote only on the feasibility and then

your overall recommendation. So, that is what these scores are reflecting. We do have all the other scores, if you want to see those as well. But this was the results of that particular survey that we had from you guys.

So, the overall recommendation ended up being 12 yes and 6 no.

CO-CHAIR ROSENTHAL: And the feasibility was 3 high, 8 medium, 6 low, and 1 indeterminate, I guess.

MS. WILBON: That=s insufficient.

CO-CHAIR ROSENTHAL: Insufficient.

So, does that give people sufficient enough information to determine whether or not we want to re-recommend it, now knowing that it is only actual prices and not the Aeither/or@?

Paul?

DR. BARNETT: I=m just confused because this is not in the report, right?

MS. WILBON: These measures? No, we didn=t put any of the Ingenix measures in

the report for this reason, this discussion right here.

CO-CHAIR ROSENTHAL: Until this question got determined --

MS. WILBON: Right.

CO-CHAIR ROSENTHAL: -- that=s why it=s not in the report.

MS. WILBON: Right.

CO-CHAIR ROSENTHAL: Once we either affirm it, if we affirm it again, it will go in the report; if we say, no, we don=t like it because it is prices only, it would not go on the report.

MS. WILBON: Well, all measures go in the report. They would just be framed as such.

CO-CHAIR ROSENTHAL: But it would go in as a negative vote, right.

MS. WILBON: Yes.

CO-CHAIR ROSENTHAL: So, is there a motion in relationship -- and I would prefer, if it is okay, we do these one at a

time -- is there a motion in relationship to 1599?

MR. AMIN: There is a degree of variability on the overall recommendation.

CO-CHAIR ROSENTHAL: Well, for example, on coronary artery disease we voted 8 yes and 10 no. And I guess the question would be, does it change anybody=s mind overall? I guess that is you are purporting to get? That=s what you are saying.

So, we could do these all at once. Does it make any difference or does it not make any difference?

DR. RUDOLPH: Do we have enough for a quorum to vote?

MS. WILBON: Yes. Twelve and two on the phone. Yes, we would have. That would be 14. Yes.

CO-CHAIR ROSENTHAL: So, I suppose the question, we can pose the question any number of ways. We could do them one at a time or we could do them in aggregate and say

the previous votes are the previous votes, and you are either voting to overturn the previous votes in aggregate, in which case we would have to do them all over again, or to reaffirm the previous votes in the notion of being consistent, and that we had both options inherent in the previous votes.

MS. WILBON: Right.

CO-CHAIR STEINWALD: Steve, would you like to make that motion?

DR. BARNETT: Could we finish -- so, there are two more measures that we didn=t review here -- just briefly what the votes were on the others?

CO-CHAIR ROSENTHAL: Yes. I=m sorry.

The congestive heart failure vote was overall recommendation, 10 yes and 8 no, and the feasibility, again, was 2 high, 8 medium, 7 low. Then, the coronary artery disease was 8 yes and 10 no, and the diabetes was 11 yes and 7 no. And interestingly, the

feasibility tracked in the same way: 2 high, 8 medium, and 8 low. So, the feasibility votes skewed low on all of these.

And to the extent, again, that there was any confusion or a clarity around this question of standardized pricing versus dollar-denominated pricing, arguably, it could change the feasibility vote.

DR. BARNETT: I=m sorry. So, that was the first two were approved and the second two were not?

CO-CHAIR ROSENTHAL: No. Three had overall recommended approvals and one did not. The three, again, the CHR vote was 10/8, yes/no. Coronary artery disease was 8 yes, 10 no. Somebody flipped, I guess. Diabetes was 11 yes and 7 no. And the non-condition-specific one was 12 yes and 6 no.

Yes, sir, Steve?

MR. PHILLIPS: So, my thought is that we would take just an overall vote. From what I am hearing, then, it seems to me that

the issue kind of cuts across all of the measures. So, I would make that motion.

CO-CHAIR ROSENTHAL: All right. So, to clarify the motion, it sounds like the motion is to keep the same votes on all four of the Ingenix measures with the information that we now know, which is they are pricing-only. That=s the motion. Okay?

Is there any further discussion? Is everybody clear on the motion?

So, if we pass the motion, then the votes that we made on these measures stand as recorded. If the vote is against this, then we have to reconsider each measure.

Now I hate to phrase it that way because that probably is going to skew the vote.

(Laughter.)

But that is what the vote would entail.

Jack?

DR. NEEDLEMAN: Tom, there are a

couple of people, you know, a couple of votes changed the endorsements on almost all these measures. So, I think the relevant question is whether anybody in the room would change their vote, based upon it only being pricing rather than both. And if there are three people in the room who would change their vote, without even asking what direction it would be, I would want to revote them.

But if nobody is going to change their vote based upon this, then I am happy to see the current vote stand.

CO-CHAIR ROSENTHAL: Well, but that ought to be, then, the basis, I guess, for people voting.

DR. NEEDLEMAN: Yes. Yes, but --

CO-CHAIR ROSENTHAL: It is, would you change your vote based on what you know? I mean that boils the question really right down to its essence.

DR. NEEDLEMAN: Yes, but a minority of the people in this room saying

they would change their vote based upon what they know would lead me to want to revote them.

CO-CHAIR ROSENTHAL: Oh, I see. I see.

DR. NEEDLEMAN: You know, three people changing their vote changes the vote --

CO-CHAIR ROSENTHAL: I see.

DR. NEEDLEMAN: -- if they all go from yes to no.

CO-CHAIR ROSENTHAL: So, as a point of order, you make a very good point of order.

(Laughter.)

Which truly meant that the motion would have to pass by a super-majority, a super-super-majority, in order to not result in the result that you describe.

We=re supposed to be chairing this thing, and I feel really sort of --

CO-CHAIR STEINWALD: We have a motion on the table. So, we should probably

vote.

CO-CHAIR ROSENTHAL: Okay. With Jack=s admonition in mind, let=s vote.

So, 1 is yes --

MS. WILBON: Well, we would just do probably a --

CO-CHAIR ROSENTHAL: A manual vote?

MS. WILBON: -- manual vote for this, yes.

CO-CHAIR ROSENTHAL: Show of hands. Show of hands.

So, all in favor of the motion?

The motion is that we would accept the votes that we took, no change, knowing what we now know, which is that Ingenix has put the thing through as a price-only measure. That=s the motion. And Jack=s point notwithstanding, that is the motion on the table.

So, a show of hands on in favor?

One, two, three, four, five, six.

CO-CHAIR STEINWALD: People on the phone?

CO-CHAIR ROSENTHAL: How do we get the people --

MS. WILBON: Jeptha, are you still there?

DR. CURTIS: Yes.

MS. WILBON: Would you like to vote now for the motion?

CO-CHAIR ROSENTHAL: He=s got to because there=s no mechanical voting.

DR. CURTIS: Yes, I would not vote in favor.

CO-CHAIR ROSENTHAL: Okay.

MS. WILBON: Not vote -- okay.

CO-CHAIR ROSENTHAL: All right. Anybody else on the phone voting?

MS. WILBON: Are there any other Steering Committee Members on the phone besides Jeptha?

DR. HALM: Yes, Ethan.

MS. WILBON: Oh, hi, Ethan. Have

you been listening to the discussion?

DR. HALM: Yes, I wouldn=t change my mind.

CO-CHAIR ROSENTHAL: He would not.

MS. WILBON: You would not change your mind?

CO-CHAIR ROSENTHAL: Okay. And then, how many are against the motion?

Two, four, five, six, seven.

They split the vote.

So, how many were -- you=ll have to tabulate the vote again. Did you count?

Seven to seven. So, the motion does not carry, which suggests to me that, in light of what Jack had said anyway, that it means we have got to go back and consider these.

The floor is open.

DR. BARNETT: I would like to move that we reopen for discussion.

CO-CHAIR ROSENTHAL: Okay. Reopen the discussion of each one?

DR. BARNETT: Individual measures.

CO-CHAIR ROSENTHAL: Individually. Okay.

And I hear a second.

Any further discussion of this?

(No response.)

What do we do if this one comes out seven to seven?

I would say, as a point of order, we have to because -- okay, well, then let=s just --

DR. BURSTIN: I was just going to also point out that you do have an option of putting something forward as a recommendation without a consensus and just getting comment, just like public comment went out today. It=s not optimal, but, truly, if it is a split, it is okay. It just means we really do need public comment to help you think that through.

CO-CHAIR ROSENTHAL: All right, that=s an option.

Well, there is a motion on the

table and seconded, which is to reopen them. Any further discussion on that?

(No response.)

All in favor?

CO-CHAIR STEINWALD: Further discussion. What do we mean by reopen?

CO-CHAIR ROSENTHAL: Reopen, I think we would revote.

CO-CHAIR STEINWALD: Yes, I know each measure, but each dimension of each measure or just --

CO-CHAIR ROSENTHAL: Well, let=s decide what it means after. We will take some executive privilege around what it means to reopen.

CO-CHAIR STEINWALD: Okay.

CO-CHAIR ROSENTHAL: Let=s have a show of hands on this one.

Can somebody count?

Opposed?

Then, we=ll get the two on the phone.

Opposed?

One, two.

Okay, and let=s get the phone votes.

MS. WILBON: And, Jeptha and Ethan, can you give your votes?

DR. HALM: I vote approval of reopening.

DR. CURTIS: And I=m okay with that. Approve.

CO-CHAIR ROSENTHAL: Okay. So, the vote was 12 to 2.

I would suggest what we mean by reopening is that we vote overall acceptability and not do each of the segments. And the one segment where this issue I think is relevant in our scoring system relates to scientific acceptability and the feasibility, the feasibility part. So, you could factor that into --

DR. BURSTIN: Usability.

CO-CHAIR ROSENTHAL: Usability,

right.

DR. BURSTIN: Yes.

CO-CHAIR ROSENTHAL: I=m sorry. Right.

So, I would suggest that we go back, and I assume maybe we could also have a suggestion that both votes be kept for the report. In the discussion, that there were two votes around this one --

MS. WILBON: Yes. Yes, we can do that.

CO-CHAIR ROSENTHAL: -- for the sake of completeness. It certainly seems to be the order of the day, completeness.

(Laughter.)

Yes, completeness and transparency.

Well, let=s start with the condition-specific ones. Maybe they will be a little less contentious.

Congestive heart failure, again, the original vote was 10 yes and 8 no for

overall recommendation.

Is there any discussion on congestive heart failure in relationship to now the question that it is prices-only?

DR. DUNN: I=m sorry. This is Dan Dunn. Could I ask a question just for clarification?

CO-CHAIR ROSENTHAL: Absolutely.

DR. DUNN: So, isn=t the question that -- I think these are two different measures. I think, if the parties agree, it is the exact same clinical logic with different assumptions about how to compute resources or costs.

Would it be an indication that one or the other isn=t good enough for a measure? Like standard prices alone or actual prices alone is not good enough, and one is not valid without the other? Is that the point? So, that means if it is just a standard-pricing-only measure, that is not enough. If it is an actual pricing measure, that is not good

enough, that both need to be available?

MS. WILBON: Dan, can you repeat your question, please? Your voice is a little muffled or something. We=ll check audio on our end, but I don=t know if you=re on a speaker.

DR. DUNN: Now is this better? Am I more clear? Hello?

MS. TURBYVILLE: We think so. Say a few more words, and let us see if it is clearer.

DR. DUNN: I=ll switch. Is this better?

MS. WILBON: Yes.

DR. DUNN: Okay. I will try to speak up. I apologize.

Shall I start from the beginning or did any of that get picked up?

MS. WILBON: Yes, start from the beginning. Sorry.

DR. DUNN: Yes, I am sorry.

I think what I am hearing is that,

and I consider it this way, that standard prices versus actual prices, there=s two different measures for each one of these considerations, and they both have exactly the same clinical logic, but different assumptions on how the resources are measured. And if that is the case, is the question that, unless you have both actual and standard, that the measure isn=t sufficient? Meaning that if you just have standard prices for a measure, that is not sufficient. If you have actual prices for the measure, that=s not sufficient, even though both could be valid, but you would have to have both for the measure to be considered? Is that the point here?

CO-CHAIR ROSENTHAL: Yes, I think that was a general consensus in the room. Well, consensus may be too strong. There were at least several people in the room who viewed them as a kind of matched pair, that you needed both for the full robustness of what they might be measuring. I suspect there

might still be some people who might either in favor -- well, there clearly were people in favor regardless and there were people against regardless, but there were at least a few people who were more inclined to be supportive if, in fact, both full pricing and the standardized pricing were a matched set. Is that a fair answer?

I am getting head-noddings around that.

DR. DUNN: Okay. Thank you.

CO-CHAIR STEINWALD: So, when we did HealthPartners, and HealthPartners originally submitted two measures as one, and we said they had to be split apart, and then we evaluated each measure independent. My recollection is that there was no co-dependency; there was no real way to factor in co-dependency in going through the process of measuring importance, and so forth.

I guess, for me, the only way that a prices-only measure or even a standardized-

prices-only measure would affect the scoring would come in usability. Because I think it is pretty clear that we established through this discussion that both a standardized pricing methodology has certain uses and an actual prices has certain uses, and they don=t necessarily overlap. You would use one for some purposes and use another for other purposes. Therefore, either one by itself has maybe less usability than a paired set.

And yet, when we went through the HealthPartners evaluation, we were evaluating each one independently. So, I can=t for myself find a logic that says, if the measure is useful for some purposes, a logic that says it is not enough to take it over the threshold unless there is another measure also independently evaluated sitting next to it. So, my logic is, especially given the process that we went through with HealthPartners, that the measure has to be evaluated independently.

***CO-CHAIR ROSENTHAL:** But we are.*

Maybe I was overchanneling Jack, but Jack had made that case.

CO-CHAIR STEINWALD: Right.

CO-CHAIR ROSENTHAL: So, one person had that feeling, anyway.

Barbara?

DR. RUDOLPH: Yes, I would speak to not making a requirement for pairing because different end-users, some will have access to pricing information, the actual costs; others will not, and they will be able to use the standardized pricing. So, I would really suggest that we not require them to be paired because in that case, then, you would have to have the actual pricing information to use the measure.

CO-CHAIR ROSENTHAL: Well, in point of reference, it is a moot question. I mean it is interesting that it was posed, but it is not a question on the table. The only question on the table is the approval of the congestive heart failure measure under the

conditions proposed, which is cost, dollar-denominated cost. I think it was a more theoretical question posed and attempted to be answered and discussed.

Got it. I got it.

(Laughter.)

DR. REDFEARN: But it seems to me that, no matter how the measure is proposed, you could choose to do something different if you wanted to do it. There is nothing in the Ingenix measure construction that requires that you use real prices or synthetic prices. You can use either in terms of the methodology, as far as I know.

Now the issue is you are voting on a measure as defined. I understand that. But it seems to me you could switch that denomination of how you denominate, either utilization or cost, you could switch that, and the method, you could just pop it right in, and it would be you could do it either way.

CO-CHAIR ROSENTHAL: I=m not sure I follow what you mean. Who could do it either way? The only NQF-endorsed measure would be dollar-denominated prices.

DR. REDFEARN: Well, that=s what I=m saying, but you could say I have the pricing methodology; I would also like to look at it from the point of view of synthetic pricing. You can do that on your own.

CO-CHAIR ROSENTHAL: Somebody could do it.

DR. REDFEARN: Somebody could do that, yes.

CO-CHAIR ROSENTHAL: Somebody could just do it.

DR. REDFEARN: Yes.

CO-CHAIR ROSENTHAL: Okay. Okay. All right, that=s fair.

Other discussion?

(No response.)

Are people okay with the notion that what we are voting on, when we vote now,

is the overall acceptability question and not going back through each segment? Because I think the point made is that, really, this decision only really affects the usability question mostly. Are people okay with that?

Are people ready to vote?

So, are we going to do the clicker thing? Help us.

MS. TURBYVILLE: I just have a quick question. I just want to make sure I=m clear. Are you saying to revote on the scientific acceptability or the overall recommendation of the measure?

CO-CHAIR ROSENTHAL: I=m suggesting overall recommendation --

MS. TURBYVILLE: Thank you.

CO-CHAIR ROSENTHAL: -- and not doing each of the four components all over again.

MS. TURBYVILLE: Okay.

CO-CHAIR ROSENTHAL: But, again, I=m open. In the spirit of trying to move it

along a little bit, but I am open if people want to or if you are telling us we have to do each segment.

MS. TURBYVILLE: No, I was just clarifying because you are using the word Aacceptability@ and we were not quite in agreement --

CO-CHAIR ROSENTHAL: Okay. Okay.

MS. TURBYVILLE: -- if you meant recommendation or scientific. So, the recommendation is fine.

CO-CHAIR ROSENTHAL: Are people clear what we=re doing? Overall recommendation.

And again, on this one, on CHF, the last time, the vote was 10 yes, 8 no. We don=t have 18 people voting. We will have 14 voting, and we will see what the vote is.

Are we going to use the clickers? So, remind us again of how to do this. And where do we point the thing?

(Laughter.)

Point it at Sarah.

MS. WILBON: Everyone point at Sarah. When she starts the voting, you will have 60 seconds to vote. It will collect your votes and will project it on the screen and read the results.

For Jeptha and Ethan, if you are still there, we will just have you --

CO-CHAIR ROSENTHAL: Can they whisper it in to Sarah since it is not exactly an open vote?

(Laughter.)

Can they whisper it in her ear and she can tabulate them?

MS. WILBON: Yes, we will just have you guys give a yes-or-no vote over the phone. Okay?

CO-CHAIR ROSENTHAL: Right. And it=s 1, yes; 2, no; 3, abstain. Okay?

This is actual pricing-only, is the proposal for Ingenix congestive heart failure.

Okay. So, let=s vote.

Are you ready? Sarah, are you working with us?

It was a little slow to start the last time. Patience will be rewarded.

MS. WILBON: We did actually test it before.

CO-CHAIR ROSENTHAL: All right. No, remember, patience will be rewarded, Helen. Remember, the last time it became fun.

(Laughter.)

The same voting rules. One is yes; 2 is no; 3 is abstain. Let=s just try it.

(Whereupon, a vote was taken.)

CO-CHAIR ROSENTHAL: Is it tabulating scores?

MS. WILBON: Yes.

CO-CHAIR ROSENTHAL: Okay.

MS. FANTA: Okay. So, 1 is yes. So, we have 5 yeses and 7 noes.

MS. WILBON: And then, right. So,

Jeptha and Ethan, are you still there?

DR. CURTIS: This is Jeptha. I vote yes.

DR. HALM: Ethan, no.

MS. FANTA: So, we have 6 yeses and 8 noes.

CO-CHAIR ROSENTHAL: Okay. Thank you.

So, next for consideration is the coronary artery disease Ingenix measure. And just for recollection, the previous vote was 8 yes and 10 no.

So, this would be open for discussion. And we would be voting, again, overall recommendation.

So, is there discussion about the coronary artery disease? The same issues, not any different.

(No response.)

I think that silence means yes.

Are people prepared to vote on the coronary artery disease measure? I=m sensing

yes.

So, we will do the same voting. It will be 1, yes; 2, no; 3, abstain. This is Ingenix 1594, coronary artery disease.

MS. WILBON: She has got to start the timer. One second.

CO-CHAIR ROSENTHAL: Okay. Hold on. Our patience is going to be tested here.

MS. WILBON: I know.

(Laughter.)

(Whereupon, a vote was taken.)

CO-CHAIR ROSENTHAL: Okay?

MS. WILBON: One yes; 2 no; 3 abstain.

CO-CHAIR ROSENTHAL: Right. Ignore what=s on the slide, other than the timer.

MS. WILBON: Four yes and 8 no.

And then, Jeptha and Ethan?

DR. CURTIS: Jeptha, yes.

DR. HALM: Ethan, no.

MS. FANTA: So, it=s 5 yes and 9

no.

CO-CHAIR ROSENTHAL: All right. And now we will consider Ingenix 1595, which is diabetes, which the previous vote was 11 yes and 7 no.

And this is open for discussion.

(No response.)

Hearing none, and assuming that the issues are largely the same, I would say we should proceed with a vote.

Are you ready, Sarah?

So, the vote will be 1, yes; 2, no, and 3, abstain.

And is the timer on? The timer is on.

(Whereupon, a vote was taken.)

Okay.

MS. WILBON: Yes, so it=s 6 yes, 6 no.

And then, Jeptha and Ethan?

DR. CURTIS: Jeptha, yes.

DR. HALM: No.

CO-CHAIR ROSENTHAL: Okay. Seven to 7. Crystal clarity on the part of the group. Well, it is crystal clear; we are evenly divided.

All right. The last measure, then, for consideration is the non-condition-specific one, which, again, as I recall, is the total cost of care, which is again the one we had the big discussion with the HealthPartners people over their non-condition-specific one. But the one on Ingenix, the vote on that one was 12 yes and 6 no. And now we would be voting on it in relationship only to the pricing-only component.

So, is there any discussion on this?

(No response.)

Hearing none, Sarah, are you ready?

Okay, 1, yes; 2, no; 3, abstain.

(Whereupon, a vote was taken.)

CO-CHAIR ROSENTHAL: Why don=t you get their votes before you announce it?

MS. WILBON: Okay. So, Jeptha and Ethan?

DR. CURTIS: Jeptha, yes again.

DR. HALM: No.

MS. WILBON: So, that=s 5 yes, 9 no.

CO-CHAIR ROSENTHAL: I think that concludes the discussion on these measures, and I think we can move on to the next agenda item.

Oh, we are ready for a break?

MS. WILBON: Yes, so let=s take a break. We=re kind of on time, huh?

CO-CHAIR ROSENTHAL: Well, we=re kind of like early.

MS. WILBON: Okay. All right.

CO-CHAIR ROSENTHAL: We=re like an hour early.

MS. WILBON: Let=s go ahead and just take an early break.

CO-CHAIR ROSENTHAL: Okay. Let=s take a 15-minute break.

We will come back and we will then consider Item 1603, which is the Ingenix ETG-based hip fracture cost-of-care measure. This will be a de novo discussion with a TAP report and the whole nine yards, like we did on all of the ones the last time.

Okay, 15 minutes.

MS. WILBON: Thank you.

(Whereupon, the foregoing matter went off the record at 10:13 a.m. and resumed at 10:35 a.m.)

CO-CHAIR ROSENTHAL: All right, let=s reconvene. Back in your chairs.

MS. WILBON: So, we are going to reconvene.

Operator, can you tell me, is Jim Weinstein or Patsi Sinnott on the phone?

THE OPERATOR: I do not have those two lines established.

MS. WILBON: So, for those in the

room, we are just trying to see if our Co-Chairs are going to be available. Otherwise, we will just kind of move forward, and we will make a list of the questions we have for them and then get them on the call when they are here.

(Pause.)

MS. WILBON: So, we are going to start with them.

CO-CHAIR ROSENTHAL: Right. We=re struggling a little bit because we are so efficient that we are an hour ahead. And the people who were expecting to be on at 11:30 to give the TAP reports, we, unfortunately, did not reach out to them at the break to see if we could get them. So, we are, I guess, reaching out to them now to see if they can join.

But we have the TAP summaries and the votes. So, I think we are now a little more familiar with interpreting what these votes mean. It just may be a little slower as

we try to do this, but I think we should move ahead.

Tomorrow morning, we may have just made one of the discussion points moot related to the harmonization issues. And the discussion of clinical logic of things I think is going to be its own kind of mindset. It gets a little more philosophical. I think to try to sort of do 20 minutes of that and then stop it and -- so, I think we will be well-served.

I think we have enough wherewithal as a group, given our experience from the last meeting and understanding now what these measures mean and what these scores, that we can, I think, interpret the TAP report. We just may be a little slower, but slower seems to me to be better than sitting and doing nothing. Right? Are we okay with that? Bruce?

CO-CHAIR STEINWALD: You know, at my age, sitting and doing nothing is always a

viable option.

(Laughter.)

But I defer to your judgment on this. Go ahead.

CO-CHAIR ROSENTHAL: All right. We could have a motion as to who would prefer to do nothing. But I don=t want to embarrass anybody on that vote, mostly myself, because I have got attention deficit disorder. So, I think I need to keep moving.

All right. So, we are going to consider, then, the hip fracture cost-of-care measure from Ingenix, No. 1603. I think our Ingenix folks are still on the phone. So, I think we would start, if you would, by having a brief description of the measure. Then, we will move into the various elements.

So, who=s on?

MS. WILBON: Ingenix folks, are you guys still there? Is Cheri or Tom still there?

MR. LYNN: Yes, this is Tom.

CO-CHAIR ROSENTHAL: Perfect. So, would you might sharing a brief description of 1603?

MR. LYNN: Yes, 1603 is an ETG-based measure around hip fracture. I am looking at capturing the cost of the condition of hip fracture as an acute disease.

It starts with the ETG methodology to gather claims to the episode of hip fracture and then goes on to evaluate the cost and some resource utilization measures around hip fracture. Of course, like the other ETG-based measures, this is a severity-adjusted measure, risk-adjusted measure.

That=s all I have.

MS. WILBON: So, just as a point of context, if you want to look at the August 5th TAP summary, that is where they discuss the 1603 measure from Ingenix.

CO-CHAIR ROSENTHAL: And would you mind just elaborating a little bit more on what the hip fracture episode of care consists

of in sort of general terms?

MR. LYNN: It uses diagnosis codes to identify episodes of hip fracture, and it is specifically hip fracture as opposed to femur fractures or pelvic fractures, and creates an episode of care that gathers all the claims around the care for that hip fracture episode.

CO-CHAIR ROSENTHAL: And how long does the episode extend?

MR. LYNN: The episode has a dynamic window. So, it extends, I believe, until there is inactivity for -- I don=t have the number right in front of me -- I think it is 90 days.

CO-CHAIR ROSENTHAL: So, it maxes out at 90 days or it can continue pass 90 days?

MR. LYNN: Every time there is an interaction between a provider and a member, a provider and a patient -- well, I shouldn=t say that -- a clinician and a patient, then

the clock restarts, and so it continues with a rolling 90 days until there is inactivity for 90 days, and then the episode closes.

CO-CHAIR ROSENTHAL: Okay. And the attribution is to whom?

MR. LYNN: I believe this rule has choices for attribution that can use either the count of encounters between a clinician and the patient or the cost of those encounters.

CO-CHAIR ROSENTHAL: Okay. Are there questions from the group about the measure itself?

(No response.)

MS. WILBON: Do you want to start with importance?

CO-CHAIR ROSENTHAL: Well, could we just -- there were three questions from the TAP that were identified. Is it worth one minute readdressing those?

MS. WILBON: Sure.

CO-CHAIR ROSENTHAL: I know you

all, Ingenix answered the questions. But there was a question about age groups with different risk factors. It looks like you answered that. Outliers at each end that were excluded.

MR. LYNN: No, outliers at the low end are excluded and at the upper end are capped.

CO-CHAIR ROSENTHAL: All right. Which is standard for their methodology, I think.

MR. LYNN: That=s correct.

CO-CHAIR ROSENTHAL: Okay. So, no other questions for the group? Yes, Steve?

MR. PHILLIPS: Yes, I just had, I guess, a general question across all the Ingenix measures that I wanted to pose to the -- I=m sorry, I missed the name.

But, in terms of defining the episode and specifically the end of the episode, I am just wondering as far as kind of the clinical input and review that the

measures go through to really get the perspective of the relevant medical societies on the decisions of when an episode ends, if you could maybe describe that a little bit.

MR. LYNN: Sure. Actually, this does have some variability amongst our measures. The chronic measures are divided into year-long segments, but the acute measures wait for a period of inactivity to call the episode complete.

We do have a panel of experts that we review these decisions with. Obviously, orthopedic surgeons, and we also have a medical advisory board that helps us in more general terms make these sorts of decisions. And that is the clinical input we receive.

CO-CHAIR ROSENTHAL: And I don=t remember the answer, I=m sorry, because I asked it five minutes ago, but I don=t remember the answer. Which is, to whom does the episode get attributed? Is it the surgeon who repairs the hip fracture? Is it the PCP

who is assigned to the patient? Is it the cardiologist who happens to consult on the case and has the majority of the E&M visits? To whom is the episode attributed?

MR. LYNN: The episode is attributed to the physician that has -- there are some options here built into the grouper. It is built into the rule. The episode can be attributed to the clinician who has the most encounters with the patient or it can be attributed to the clinician with the most dollars caring for the patient.

That is limited to a list of specialties that would be allowed to win such an episode. And I believe in this case that it is really only orthopedic surgeons that can win this episode. Or it is limited to a certain peer group.

CO-CHAIR ROSENTHAL: So, I=m sorry, the last thing you said was only orthopedic surgeons can get the episode attributed to them?

MR. LYNN: That=s correct.

CO-CHAIR ROSENTHAL: Okay.

MR. LYNN: If a cardiologist were to win the episode, it would not be included in the analysis.

CO-CHAIR ROSENTHAL: I think we have Dr. Weinstein on the phone, who chaired the TAP Committee on this.

So, Jim, we are going to start through, then, the scoring measures. You could start if you have any general comments. Otherwise, we are going to go through in sequence importance, scientific acceptability, et cetera, and you could make specific observations about each of those segments as we get to them.

DR. WEINSTEIN: Okay. Thank you.

Yes, just overall we are talking about hip fractures, this one?

CO-CHAIR ROSENTHAL: Yes, that=s correct.

DR. WEINSTEIN: Yes. I think we

were pretty explicit as a group that the limitations of this were the commercial database did not have a population of patients greater than 65 for the most part where most of these occur. And we worried that the attribution, comorbidities, and some other things related to younger patients would not be seen in this and, therefore, may make the model suspect. That was the major concern, just the focus of the age of the population, which I think is brought out in the documents several times.

But, truly, hip fractures in people less than 65 are much different than people over 65. In fact, there is some data suggesting that there has been a decade in change in the rates of these towards older people with more complicated fractures. That is from the Mayo data in their community there. They have really done a large cohort of patients over time. So, that was a major concern of the group.

And I=ll stop there.

CO-CHAIR ROSENTHAL: All right. Well, that is a helpful overview.

MS. ZIELINSKI: This is Cheri Zielinski with Ingenix. Can I just add a comment?

CO-CHAIR ROSENTHAL: Absolutely.

MS. ZIELINSKI: Thank you.

We did specify this as a commercial-based measure and not a Medicare-based measure. So, we used the commercial population.

DR. WEINSTEIN: And we discussed this, and you=re absolutely right. I did offer, through the Dartmouth Group, to actually do some of this, if you wanted to run it on a 65-plus population during the Committee meeting. But it is a limitation, so we just need to be clear.

CO-CHAIR ROSENTHAL: All right. Well, I think, then, we will consider importance. I think sticking with our theme

from the last meeting, although I do notice that the TAP, unlike every one of the measures from the last meeting where the TAPs all basically were unanimous about the importance, it looks like the TAP vote on even importance was a bit split. But, nonetheless, I think the action is going to be still in scientific acceptability, usability, and so forth.

We could have an extensive discussion about importance, if anybody would like to discuss the importance question. Otherwise, I think we would move to the vote on that.

Okay. Ashlie, it is a little hard to see.

MR. AMIN: Tom, could I just clarify one thing?

CO-CHAIR ROSENTHAL: Yes.

MR. AMIN: The TAP discussion on importance here, and I think as Dr. Weinstein has pointed out, was around whether this measure would be important to measure in a

population that is under 65. So, that could occur in the importance section.

DR. DUNN: Yes, and do you know anything about the epidemiology of the proportion of these that are really 64 and younger? I would think it is 5 or 10 percent.

CO-CHAIR ROSENTHAL: Well, it just may mean that the importance vote, it may not be unanimous as it was in each of the ones that we had the last time.

But, Taroon, I can=t read this at all. But this, I assume, is the four elements of importance from the TAP Committee. So, could you help us orient those? Or Sarah?

MS. FANTA: Sure.

CO-CHAIR ROSENTHAL: Thank you.

MS. FANTA: Yes, it is regarding high impact of care or high impact; opportunity for improvement; demonstration of resource use, problems and variation; the purpose is clearly described, and the resource use service categories are consistent with the

intent. That encompassed importance.

MS. WILBON: Oh, on the TAP ratings graph that is projected, we have got for high impact, there was 7 high, 2 moderate, and 2 low. For 1b, which is the second bar from the left, you had 5 high, 4 moderate, and 2 low. For 1c, which is the purpose is clearly described, you had 2 high, 8 moderate, and 1 low. And then, for the resource use service categories are consistent and representative of the intent, you had 4 high, 5 moderate, and 2 low.

CO-CHAIR ROSENTHAL: Thank you.

And our choice in the overall importance is yes/no. So, 1 will be yes and 2 will be no.

And, Sarah, are you ready for the vote?

MS. FANTA: I hope so. Let=s see how it goes.

CO-CHAIR ROSENTHAL: All right. Here we go.

(Whereupon, a vote was taken.)

MS. WILBON: So, especially now since the vote is split on importance, we are going to need the people, the Steering Committee Members on the phone to provide a vote on overall importance.

DR. WEINSTEIN: Yes, so this is Jim again. I think there was nobody on the Committee, at least from my recollection of the meeting, that didn=t think hip fracture wasn=t important.

And it is confounded by this age issue and the data system. It is Ingenix=s fault. They were responding to the request. But the issue is this is a different population, and it is extremely important. There is a 30 percent one-year mortality with these patients. So, it=s a big deal.

CO-CHAIR ROSENTHAL: Jim, I think we got it. I think what we are trying to do is we are voting with a little machine here in the room, and we want to be able to count the

votes of the people on the phone, yourself included. And unfortunately, since you don=t have the little machine, we have to ask you to, in effect, give a yes or no vote on importance to the staff, who will kind of incorporate that into the overall vote.

So, if we could get each of the -- I think there are now three on the phone who are Committee Members, and let=s get those votes, if we could.

So, Ashlie, the question?

MS. WILBON: So, for those Steering Committee Members on the phone, we just need a yes or a no vote.

DR. PETER: This is Doris, and I voted yes.

MS. WILBON: Oh, Doris, okay.

I don=t know who else is there. There may be some others.

Is Jeptha still there?

DR. CURTIS: Yes, I vote yes.

MS. WILBON: Okay. Ethan?

DR. HALM: Yes.

MS. WILBON: And Jim Weinstein?

DR. WEINSTEIN: Yes.

MS. WILBON: Okay. Are there any other Steering Committee Members who were able to dial in?

(No response.)

Okay. Okay, thank you.

MS. FANTA: Okay. So, the total vote was 10 yes and 6 no.

CO-CHAIR ROSENTHAL: Okay. So, now we would move on to scientific acceptability. We start, then, with reliability, right?

So, Jim, we will turn this back over to you, then, to discuss the TAP view of the reliability, which as specified says that the measure is well-defined and precisely-specified.

DR. WEINSTEIN: I think that was fine. I don=t have a comment on that. Yes. If that is a vote, I am going to say that it

is a reliable acceptable measure issue, given the issues that we have already talked about.

CO-CHAIR ROSENTHAL: Okay. And then, the second part of the reliability is that the results are repeatable.

DR. WEINSTEIN: Yes. I mean the issues, I don=t know what documents you have in front of you, but, again, with this age population and the comorbid conditions and the issues around reliability, it was hard to tell some of that from the tables that we got.

And again, this is all a little bit undermined by the whole population issue, I am sorry to say. But I don=t want to keep repeating it, but that is the issue because it affects everything else.

CO-CHAIR ROSENTHAL: So, Jim, would you just elaborate a little bit, because the overall reliability vote from the TAP was 1 high, zero medium, and 4 low.

DR. WEINSTEIN: Yes, I don=t have that voting in front of me. So, I don=t know

what it was, but if that=s what it was.

***CO-CHAIR ROSENTHAL:** Do you have a sense of what the low was being driven by?*

***DR. WEINSTEIN:** I am just guessing the reliability, given the fact that the specific data that is missing from this population doesn=t allow it to be reliable. And if the other group members want to speak up? But it is like comorbid conditions are very different in a young population than they are in an older population.*

***CO-CHAIR ROSENTHAL:** And those comorbid conditions are not accounted for in the risk-adjusting methodology from Ingenix?*

***DR. WEINSTEIN:** Right.*

***CO-CHAIR ROSENTHAL:** That would be troublesome.*

Open for discussion around reliability.

***DR. CURTIS:** But just to clarify -- this is Jeptha -- aren=t they requesting an endorsement for use in a commercial population*

alone? So, whether or not it is appropriate to apply it to a Medicare population would seem --

CO-CHAIR ROSENTHAL: Yes, this is only a commercial population measure, under 65.

But, Jim, those other comorbid conditions that would impact outcomes, are they relevant in an under-65 population in the same way that they are relevant in an over-65 population?

DR. WEINSTEIN: Well, they would be relevant, but they don=t occur as often, obviously. Therefore, they are not variables that we would think would impact on the overall outcome or resource utilization, et cetera.

CO-CHAIR ROSENTHAL: Okay. Other questions or comments from the Committee?

(No response.)

So, I think our task, this now will be 2a, which is overall reliability,

which captures the two notions of well-defined and specified and repeatable. We would be ready to vote.

And, Sarah, would you mind giving the TAP scores on which bars and what the scores were there?

MR. AMIN: I can do that. So, 2a1 would be well-defined and precise specifications, the bar all the way to the left. It was 3 high, 5 moderate, and 2 low. And reliability testing, of 2a2, the second bar from the left, 3 high, 3 moderate, and 4 low.

CO-CHAIR ROSENTHAL: That is the only part that puzzles me a little bit, is I don=t know what the basis of the 4 lows were on this being repeatable. The measure I assume has been tested in a variety of settings? I mean that would determine whether it is repeatable. But has it been tested widely?

DR. WEINSTEIN: I=m not sure it

was tested in multiple settings.

MS. WILBON: I think it had to do with the TAP=s difficulty in understanding the information that Ingenix submitted to demonstrate reliability. I think they were just having trouble navigating, understanding, interpreting what they submitted as evidence of reliability.

DR. REDFEARN: What it says in the notes is, AThe panel questioned whether one can infer group or reliability from the table submitted by Ingenix.@ That=s the comment.

CO-CHAIR ROSENTHAL: Is the group prepared to vote on the reliability, 2a, question?

I think we are trying to clarify our recollection of the previous meeting, but I think we voted on the subsections and then we voted on overall scientific acceptability.

MS. WILBON: Yes.

CO-CHAIR ROSENTHAL: Right, Ashlie?

MS. WILBON: Right.

CO-CHAIR ROSENTHAL: I am trying to follow advice of counsel here.

MS. WILBON: Yes, we are going to vote on the overall reliability, overall validity. Even though the TAP did that as well, we also kind of want the Steering Committee=s votes on those. And then, we will have you also vote on the overall scientific acceptability, just to be consistent in the way we have been doing it for the process thus far.

CO-CHAIR ROSENTHAL: Which is what we did the last time.

MS. WILBON: Right.

CO-CHAIR ROSENTHAL: At least that is my recollection as well. But she is the boss on this one. So, we will vote on each of these in sequence.

And again, the TAP vote are the two bars farthest to the left on this. Okay? All right.

The vote here is high, moderate, low, and insufficient, correct?

MS. WILBON: Yes.

CO-CHAIR ROSENTHAL: Right, that=s the vote on this. And then, when we do overall scientific acceptability, it will be yes or no.

Yes, let=s revote on this. I=m sorry, my fault.

One is high, 2 is moderate, 3 is low, and 4 is insufficient.

(Whereupon, a vote was taken.)

Can we get the phone votes then as well.

MS. FANTA: All right. Jeptha, we are voting right now on overall reliability, high, moderate, low, or insufficient. Jeptha, are you there?

DR. CURTIS: Yes. Moderate.

MS. FANTA: Moderate, okay.

Doris Peter?

DR. PETER: Moderate.

MS. FANTA: Okay. Jim?

DR. WEINSTEIN: Moderate.

MS. FANTA: Okay. Ethan?

DR. HALM: I just said moderate. Sorry.

MS. FANTA: Okay. Thanks.

Patsi? I=m not sure if you=re there. She=s joining. Oh, sorry. Okay.

All right, then. All right. So, we have 1 high, 11 moderate, 3 low and 2 insufficient.

CO-CHAIR ROSENTHAL: All right. So, let=s now move to validity. And let=s see, there are six measures of validity. Evidence is consistent with intent, exclusions, risk adjustment, identification of statistically-meaningful differences, and multiple data sources.

So, Jim, would you mind giving the TAP discussion on validity?

DR. WEINSTEIN: This is on the 2b?

CO-CHAIR ROSENTHAL: Yes, this

would be the various 2b elements.

***DR. WEINSTEIN:** Yes. I think, again, unfortunately -- I sound like a broken record -- but the commercial population was a small number of these patients in their overall population because the incidence of this is fairly low in this commercial population. So, the panel was very concerned about the validity of this, given that fact.*

And again, we are thinking of hip fractures as a very common problem, but what we are testing here is something that is uncommon.

***CO-CHAIR ROSENTHAL:** Yes, David?*

***DR. REDFEARN:** I am looking at the counts for this. One of the questions I have is, what happened to all the votes when you look at the final overall validity? There=s only four people voting on the overall validity when you have up to nine votes on the individual components. Why didn=t people vote?*

MS. WILBON: *That might actually be a typo. I=m not really sure. We will have to go back and check. Oh, we did this one on the phone.*

CO-CHAIR ROSENTHAL: *Yes, but his point is that the subsections all have fairly high numbers. So, if you just took 2b2, for example, there were no high, 3 medium, and 7 low. So, there were 10 voting people. And then, when you get to overall voting, there is only four votes. Were they done asynchronously?*

MS. WILBON: *Yes. Well, I have to double-check that. I think there is probably a typo in here somewhere, to be honest with you.*

CO-CHAIR ROSENTHAL: *All right. Well, can you identify --*

MS. WILBON: *We will double-check that.*

CO-CHAIR ROSENTHAL: *All right. They will find out whether this is a typo or*

what the cause of that is. Good pickup.

Open for questions.

DR. RUDOLPH: If this measure was named something that actually described the population, would the TAP have considered this to be a valid measure of the commercial population, not invalid because it is not measuring something else? In other words, if, in fact, it was ETG-based hip fracture resource use measure for commercial population, would that have changed the vote?

DR. WEINSTEIN: I think people might have seen it differently. But the issue would still be the same because at that point you are getting into whether this is an important measure, and we would say in that younger population it wouldn=t be.

CO-CHAIR ROSENTHAL: Can I ask the question slightly differently? If in commercial populations this is an uncommon event, is the measuring, are the comparisons reliable, given the small numbers that are

likely to be involved, particularly -- let=s make it up -- that you have got a medical group that has got five orthopedic surgeons and a commercial population of 100,000. Each sees two hip fractures -- I=m making it up totally -- two hip fractures per year. Are you going to end up with meaningful differences? Was that a factor in the TAP=s thinking on it?

DR. WEINSTEIN: I would say no.

DR. SINNOTT: This is Patsi Sinnott. I=m sorry I=m late.

I was a member of the TAP. I just wanted to add -- I think that=s Jim, right?

DR. WEINSTEIN: Yes.

DR. SINNOTT: Jim=s comments.

The issue about reliability overall for the ETG product is that they produced no information that compares scoring or attribution of episodes over time. So, one of the big issues in measuring resource use for a population of physicians is that you

would expect physician performance to be fairly consistent, and that the, quote, Ascores@ or the cost attribution, or whatever, should be due to the physician practice, not to variation in patient population, and that you would want to be controlling for variation in patient population.

So, what they showed us in terms of reliability of the grouper function was that, if they took the data and grouped and then assigned to a provider, at any one time the scores ended up approximately the same. But they didn=t show us that, if they repeated it in multiple sets of the data, that the scoring was free from abnormal severity or unusual severity. I hope that=s clear, what I am trying to say. It is that the grouper function was not tested and not reported on.

***CO-CHAIR ROSENTHAL:** Okay. In multiple settings over multiple times.*

***DR. SINNOTT:** Right.*

***CO-CHAIR ROSENTHAL:** Okay. And I*

am going to re-ask my question again, though. So, everybody got her input from the TAP? I want to ask my question again because 25b, at least in the notes we have, it says the TAP discussion, AThere was a discussion regarding the relative cost-of-care ratio and a question about what numbers represent statistically-significant differences, and a suggestion that the underlying variance of episode cost in the total number of cases@ -- and this ended up scoring six out of, well, six, seven, eight, nine out of the ten voted low or indeterminate on the ability to detect statistically-meaningful differences.

So, can somebody comment either from the TAP about what the thinking was there or from Ingenix about how to answer that?

DR. WEINSTEIN: How to answer? I=m sorry. How to answer the --

CO-CHAIR ROSENTHAL: Well, 25b says, AIdentification of statistically-significant and meaningful differences,@ which

I=m assuming means that, if you apply this measure to Group A or Group B or Doctor A and Doctor B, that this thing will detect statistically-meaningful differences accurately. And the TAP vote was --

DR. WEINSTEIN: We didn=t think so.

CO-CHAIR ROSENTHAL: Okay. All right. Well, that seems to me the essence. I am trying to move it along here, folks. It seems like to me sort of the essence of reliability and validity, but I=m trying to make sure that we either get an answer from the TAP as to what the thinking was or an answer from Ingenix that satisfies this group to the contrary, so that we can have an informed decisionmaking process here.

MR. AMIN: Tom, do you think Doctor A versus Doctor B is like too tough a standard? I mean Region A versus Region B or Delivery System A versus Delivery System B, that might be more reasonable.

CO-CHAIR ROSENTHAL: ***Well, we=ll get to the attribution and its importance later.*** ***But the attribution certainly is relevant to how statistically-significant they interconnect.***

So, maybe we can ask the question -- I thought we asked and answered it -- to whom is this attributed? ***And it can be attributed in the rule set that is applied by Ingenix to individual orthopedic surgeons.*** ***So, it is a pretty high hurdle.*** ***And consequently, I think that is relevant in one=s decisionmaking around whether one is going to consider this to be statistically-accurate or not.***

MR. AMIN: ***Tom, can I offer one piece of clarification?***

CO-CHAIR ROSENTHAL: ***Yes.***

MR. AMIN: ***In order to separate the level of measurement or level of analysis and the attribution approach, the point is still valid in that this measure is submitted***

for the level of measurement, including at the individual provider level, which your point is clearly valid.

And then, it is also attributed at the group practice level, at the facility level, health plan, and further up, but --

CO-CHAIR ROSENTHAL: Well, we have to take it as it is written. If it were written that said it would only be attributed at the health plan level or at the group level, then that would be the basis under which we should consider statistical validity. If it is down to the individual physician level, then it seems to me that it would need to be accurate at the individual physician level in order to consider it statistically-reliable, unless I am missing some aspect of the way we should be thinking about this.

But, again, I am open for discussion.

DR. RUDOLPH: I am just thinking about in some places, like Wisconsin, there

are really large practice groups, and a practice group might be able to look at individual physicians and could be statistically-significant when you are looking at, you know, like 10 orthopedic surgeons, or whatever, or 20 in the group.

So, I think their response was that it would depend on, statistical-significance would depend on the numbers of total cases that there were and dependent on the confidence interval that you wanted to use, whether it was the 95th percentile or 90th whatever.

So, I don=t know, if they were to have to prove this to us, how would they do that?

DR. REDFEARN: I think, in general, the way they have answered this question is to say you can=t look at the numbers alone; you would have to apply a statistical measure. And they are suggesting you use confidence intervals.

I think the way this plays out is that confidence intervals are sensitive to how big a sample you=re looking at and how variable the underlying data is. And they go together in terms of where it falls in the confidence interval. So, they are just answering it.

So, it is a legitimate question to say, if this is so rare in the population that you are looking at, you are going to have a very small sample size. The end result will be you will say you don=t know, and you won=t be able to do that evaluation. It just depends on the data.

CO-CHAIR ROSENTHAL: Well, that=s right, and that=s why either the TAP asked the question and either had it answered or not or we can ask it again.

In the settings where it has been tested, what does it show? Does it discriminate or doesn=t it discriminate?

DR. LEE: Well, I=m sure it

discriminates. Whether it gives you actually useful information is another question.

I mean I actually think I am not too agitated about it because I actually think people have common sense and they can recognize when a measure is being used in a ridiculous situation and when it is not.

But I do think what we are seeing is that measures don=t exist in a vacuum, and it does matter the size of the patient sample. And when you get down to an individual doctor level, most of these are going to end up getting low votes from people who are being thoughtful.

That said, I don=t think that means the measure is bad. I think that the measure can be very useful at a bigger scale, at a higher level.

CO-CHAIR ROSENTHAL: All right. Other discussion, then, about validity?

(No response.)

Hearing none, I think it is time,

then, to vote. And again, now I will reiterate that the scoring will be 1, high; 2, moderate; 3, low, and 4, insufficient.

And again, either Taroon or Sarah, or whoever is going to do it, if you would reiterate the TAP scores, and not that we have to be slavishly adherent to the TAP scores, but the TAP folks did spend a day looking at this in more detail than we do, and it is there for our consideration.

DR. HALM: Before we get to that, can someone just remind us what risk adjustment was done if there are no comorbidities in the risk adjustment? Because that was the individual criteria that looked the worst.

DR. WEINSTEIN: There are comorbidities in the criteria. I think the point was that they would be different comorbidities if you looked at the over-65 population. We used the morbidity --

CO-CHAIR ROSENTHAL: All right.

So, the answer is there were no comorbidities factored into the under-65 population, with the logic being if you were taking a 65-and-over population, you would certainly have them. That is what I am here.

DR. WEINSTEIN: Well, it is not being actually approved --

CO-CHAIR ROSENTHAL: Oh, I misheard then. I=m sorry. So, what are they? I didn=t hear what they were. I think the question was, what were they?

DR. WEINSTEIN: What are the comorbidities? Is that the question?

CO-CHAIR ROSENTHAL: I think the question is, what were the comorbidities that were factored in generally?

DR. HALM: Don=t worry about that. I just wanted to make sure there were a lot of them as a class.

DR. WEINSTEIN: Yes, there=s a lot of comorbidities.

CO-CHAIR ROSENTHAL: All right.

I=m sorry. I misheard. I misheard.

Other questions in relationship to this?

MR. AMIN: Tom, I would also offer that Carlos, our statistical consultant, is on the phone, if you have any questions.

CO-CHAIR ROSENTHAL: Oh, absolutely. So, Carlos, would you mind taking a moment, then, to comment on the statistics on this? And we appreciate your being available to give us your opinion.

MR. ALZOLA: Okay. Thank you.

Yes, one of the issues that became clear to me after hearing this discussion is that --

CO-CHAIR STEINWALD: We can=t hear you, Carlos.

MR. ALZOLA: Okay. I=m sorry.

One of the issues that became clear to be after hearing your discussion is that the sample sizes are likely to be small if we try to apply the measure at the

individual physician level, especially if we are considering only the commercial population. So, that does not mean that it won=t be useful at a higher level, as it was mentioned.

In terms of the comorbidities, I am trying to open my data sheet, but I do remember that there were a lot of comorbidities used in the model. I can=t tell you which ones right now.

What else? In terms of reliability, I thought that the measure was reliable in terms of their ability to be repeatable. One of the things they did is tested the measure and they developed the data using two completely different approaches to see if they arrived at the same dataset, and they did. The two datasets match in 99 percent of the cases.

And they also look at repeatability in looking at the nine different HCOs. And, yes, of course, there was

variability because they were not using standard prices, so there would be the natural variability that you would expect from being in different markets and different agreements with insurers. But I thought they were reasonable. The variability I saw was reasonable.

CO-CHAIR ROSENTHAL: All right. Thank you.

And, Taroon, are you going to tell us, remind us again of the TAP scores here and which bars are which?

MR. AMIN: So, we will just give you, for 2b1, it was 5 low. For 2b2 -- oh, so, there was a question on the end, the difference in the number of respondents.

So, we may have some issue with the SurveyMonkey, which is why the data on your sheets may not be correct. So, I am presenting the actual correct data from SurveyMonkey, just to make sure that we all have the full information.

CO-CHAIR ROSENTHAL: Okay. So, what are we looking at on this slide there?

MS. FANTA: This is just overall --

MR. AMIN: Yes.

MS. FANTA: It is some criteria and what rolled up to that.

MR. AMIN: Of validity.

CO-CHAIR ROSENTHAL: Okay.

MR. AMIN: Can you go back to the specifics?

So, 2b1 is specifications consistent with resource -- honestly, I can=t read it myself, 2b1.

CO-CHAIR ROSENTHAL: Who=s got it on a slide there? Come on, somebody with a computer, and just tell us what it says.

MS. FANTA: Overall validity encompasses the specifications that are consistent with resource use and cost problem. The validity testing, the risk adjustment, and identification of statistically-

significant are meaningful differences.

DR. PETER: And exclusions.

CO-CHAIR ROSENTHAL: There are five bars up there that I can=t see what they are that we have not done. Which of the five -- can somebody just point out what=s what, just so we all are on the same page?

MR. AMIN: Okay. So, let=s do this: 2b1 is specifications consistent with resource use and cost problem.

MS. DORIAN: And that was 5 low.

MR. AMIN: 2b2, validity testing --

MS. DORIAN: Four low, 1 medium.

MR. AMIN: 2b3, exclusions.

MS. DORIAN: Four low, 1 medium.

MR. AMIN: 2b4, risk adjustment.

MS. DORIAN: Four low, 1 insufficient.

CO-CHAIR ROSENTHAL: Yes, those are the bars. So, that was just the fourth bar, the second one from the right, correct?

DR. BARNETT: What they are reading is different because the bars are in error. What is in the report is in error.

CO-CHAIR ROSENTHAL: Oh, okay.

DR. BARNETT: So, they are reading the results off the original source.

CO-CHAIR ROSENTHAL: I=m the only one that didn=t understand that. Thank you for explaining it.

(Laughter.)

MR. AMIN: I apologize for the confusion.

And 2b5, identification of statistically-significant and meaningful differences.

MS. DORIAN: Four low, 1 insufficient.

MR. AMIN: Is there any of the subcriteria that you --

CO-CHAIR ROSENTHAL: Okay. So, that=s now clear. And so, then, they would have had a vote on overall validity.

DR. RUDOLPH: *Are you saying there is only a total of five people on the TAP? That was it?*

MS. WILBON: *There were only six, I think, but we did this on a call, and we had them go into the SurveyMonkey after the call.*

DR. RUDOLPH: *Okay.*

MS. WILBON: *So, I think there were like six people on the call. So, five of the six people responded to the survey on the call.*

CO-CHAIR ROSENTHAL: *And I think this is about the size of the votes that we had on the TAPs from the last meeting, right?*

MS. WILBON: *Yes. This was a smaller TAP because we only had like four measures.*

CO-CHAIR ROSENTHAL: *So, what id the TAP vote on overall validity?*

MS. DORIAN: *That was 3 low and 1 medium.*

CO-CHAIR ROSENTHAL: *Okay. So,*

the discrepancy on the piece of paper is that there were, in fact, a small number of people voting on each one of the measures, and that the submeasure votes are typos on the paper that we are looking at. Okay.

All right. So, with all of that clarification, then are we prepared to vote on overall validity? And it looks like the answer is yes. And the voting here will be 1, high; 2, moderate; 3, low, and 4, insufficient.

So, is everybody clear, including me? I=ll answer for me. I think I finally get it.

So, Sarah, can we do that?

(Whereupon, a vote was taken.)

MS. FANTA: Okay. And for those of you on the phone, again, we are voting on overall validity, high, moderate, low, or insufficient.

Jeptha?

DR. CURTIS: Low.

MS. FANTA: Okay. Doris?

DR. PETER: Moderate.

MS. FANTA: Jim, are you still there?

DR. WEINSTEIN: Yes.

MS. FANTA: Okay.

DR. WEINSTEIN: Moderate.

MS. FANTA: Okay. And Ethan?

DR. HALM: Moderate.

MS. FANTA: Okay. Thank you.

So, we have zero high, 6 moderate, and 10 low, and zero insufficient.

CO-CHAIR ROSENTHAL: All right. Now if I could get clarification, do we need to vote on 2c, the stratification for disparities? I don=t remember doing that last time.

MS. FANTA: No. Just overall.

CO-CHAIR ROSENTHAL: So, now we would move to overall scientific acceptability, which would factor in all of these elements. And this one is 1, yes, and

2, no.

Is there any further discussion on the general scientific acceptability?

(No response.)

All right, hearing none, Sarah, do you want to start the clock?

(Whereupon, a vote was taken.)

MS. FANTA: So, real quick, for those of you on the phone --

CO-CHAIR ROSENTHAL: Let=s get the phone vote --

MS. FANTA: Right. Yes, for everyone on the phone --

CO-CHAIR ROSENTHAL: -- before we read votes.

MS. FANTA: Yes. For everyone on the phone, if you could vote on overall scientific acceptability, either yes or no.

Jeptha?

DR. CURTIS: Yes.

MS. FANTA: Okay. Doris?

DR. PETER: Yes.

MS. FANTA: Okay. Jim?

DR. WEINSTEIN: Yes.

MS. FANTA: Okay. And Ethan?

DR. HALM: Yes, reluctantly.

(Laughter.)

MS. FANTA: Okay. So, we have 7 yes and 10 no.

CO-CHAIR ROSENTHAL: Now, if I understand our rule set, thus endeth the conversation, 7 yes, 10 no, for scientific acceptability.

MS. WILBON: Remember, we are kind of following, we are being consistent with how we have done it before and allowing the Committee to vote on overall scientific acceptability.

CO-CHAIR ROSENTHAL: I think we did it on the other one because the vote was like 9 to 10 or something. Or I don=t know. We can do it any way the group wants to do it.

Helen, what is your advice?

DR. BURSTIN: And the individual

breakdown by reliability and validity?

MS. WILBON: Overall reliability --

DR. BURSTIN: Use your microphone, Ashlie. Sorry.

MS. WILBON: All right. Overall reliability was 1 high, 11 moderate, 3 low, and overall validity was 6 moderate and 10 low. So, validity would really strike it out.

DR. BURSTIN: Validity went down. So, essentially, it=s down. Right. Yes, agree.

CO-CHAIR ROSENTHAL: In my opinion, it wasn=t like the other one where it was really split and we moved on. And besides, it was the same measure where we had accepted the other one and disapproved the one. So, I think we are done, right, Helen?

DR. BURSTIN: Right.

MS. WILBON: So, the next measure, which is also a bone joint measure, is 1609, which is the ETG-based hip and knee

replacement cost-of-care measure. And the votes on these are definitely correct because it happened at the in-person meeting and we captured those correctly. So, we shouldn=t have those same issues.

CO-CHAIR ROSENTHAL: All right. So, if we could ask, if Ingenix wouldn=t mind giving us a brief synopsis? And then, Jim, we will ask you to give us a little synopsis from the TAP. And then, we will move into the segments on 1609.

MR. LYNN: This rule was based on a slightly different technology. We used the technology called procedure episode groups, which runs on top of the episode treatment group process.

We identify what we call the anchor procedure, which is the hip or the knee replacement. We look at a fixed time window in a short period and long time period around that anchor. We basically take all claims in a short time period that have consistent

diagnostic information on them. And then, in the further time windows we require the diagnosis information as well as specific procedure codes that are known to be part of the sequence of care for hip and knee replacement or a potential complication.

And that is the overview of this group. The rest of it is relatively the same as the other rules where it gathers those to some entity, whether it be a group or a physician or a health plan, and does a similar metrics going forward.

CO-CHAIR ROSENTHAL: Jim, would you just give us a quick overview for the TAP? And then, we will get into the various elements.

DR. WEINSTEIN: Yes. We still run into the similar issues around the commercial, but less so. But it is still an issue.

But I think that the claims data was grouped into service categories to better identify where utilization was high and low

and where the majority of cost and the treatment was. And they included specialty services, inpatient services, radiology service, et cetera.

And once the data was grouped, how to apply the cost metric to the utilization data was done, but there was no recommendation for a clear method to me in understanding this. And one of the questions that came up was whether or not the data could be customized if there were differences in the logic groupings. These are overall issues.

It seemed useful overall. It didn=t address the issue of specific resource utilization within a procedure or an E&M visit; i.e., the type of provider or non-billable activities.

So, those are some of the comments.

CO-CHAIR ROSENTHAL: Are there questions from the group about any general issues? Is everybody clear on what it is?

And then, we can move into the various elements. Any questions at this point?

Barbara?

DR. RUDOLPH: No.

CO-CHAIR ROSENTHAL: All right. I think, then, let=s consider the importance question. Is there any discussion about importance?

Jack?

DR. NEEDLEMAN: I=m looking at the HCUP data on total knees and total hips for 2009, and it looks like nearly half of total knees and nearly half of total hips are in patients under 65.

CO-CHAIR ROSENTHAL: Right. So, the critique that was relevant, apparently relevant, in the fractures --

DR. NEEDLEMAN: In the fracture, it was about 12 percent.

CO-CHAIR ROSENTHAL: You=re saying that this is a 50/50 and, therefore, the fact is that this is a more relevant condition in

the commercial population that is being measured here?

DR. NEEDLEMAN: Yes.

CO-CHAIR ROSENTHAL: Okay. Any other discussion around importance?

(No response.)

So, I think we=ll call the question on this one. And the importance here, the vote is 1, yes; 2, no. It is either important or not important. So let=s go ahead with this vote.

(Whereupon, a vote was taken.)

MS. FANTA: Okay. And for everyone on the phone, we will go ahead and vote on importance, yes or no.

Jeptha?

DR. CURTIS: Yes.

MS. FANTA: Doris?

DR. PETER: Yes.

MS. FANTA: Jim?

DR. WEINSTEIN: Yes.

MS. FANTA: Ethan?

DR. HALM: Yes.

MS. FANTA: So, 17 yes.

CO-CHAIR ROSENTHAL: There was one more. Was it Jim?

MS. FANTA: Jim voted.

CO-CHAIR ROSENTHAL: Oh, I=m sorry.

MS. FANTA: That=s okay.

CO-CHAIR ROSENTHAL: How about Patsi? Is she still on --

MS. FANTA: She=s not on the Steering Committee.

CO-CHAIR ROSENTHAL: Oh, she=s not on it.

DR. SINNOTT: She=s here, but she=s not on the Steering Committee.

(Laughter.)

CO-CHAIR ROSENTHAL: Oh, okay, she is just part of the TAP. I=m so stupid.

Well, we have unanimity at last.

MS. FANTA: Seventeen overall, yes.

CO-CHAIR ROSENTHAL: Everybody believes that this is important. Okay. Excellent.

So, let=s now move to the various aspects of scientific acceptability, and the first portion of this will be 2a, reliability, the same discussion as last time.

So, Jim, again, would you give us the TAP thinking on this?

DR. WEINSTEIN: Yes, I think this is true of a lot of databases, but right and left is a problem and it is an important issue.

One of the issues in things like hip replacement and knee replacement is patient preferences. So, you might have, as was stated, I think somebody stated some dataset suggested that half of these patients are done under 65. One would wonder about the incidence of those or the rates of those procedures in those populations being good or bad. And so, patient preferences, given good

information, what would they be, and that is sort of the topic that we elucidated in the summary there.

I think we thought, as I mentioned in my opening comments, that some of the logic and specific codes could have been clearer for us, but those are the comments.

CO-CHAIR ROSENTHAL: Questions from the group? Tom?

DR. LEE: A comment and a question. I mean I think that compared to almost everything else we do in medicine, there is like more homogeneity. We can find lots of reasons, you know, worry about risk adjustment, but like risk adjustment is almost a bigger issue than virtually everything else that we look at.

Now, that said, Jim, in the patient preference thing, how big of an issue is doing two knees at a time versus one knee at a time? It seems like it is something that paralyzes a lot of us in our interaction with

patients, but is it really like not that big of an issue, not that common a topic enough to be worrying about here?

DR. WEINSTEIN: It is a very interesting question because I have been looking at lots of different databases. And there are some institutions, as you know, that do simultaneous bilateral knees. There are some institutions that do them two different settings. There are some institutions that do one right and left separately in the same session.

So, the incidence of bilaterality is not insignificant. So, the counting issues become important, and the way you get the rates becomes important.

So, it is not as uncommon as I thought it was, but I would say most orthopedic surgeons would say you probably shouldn=t be doing them concomitantly because of complication rates, but there are some institutions that do.

CO-CHAIR ROSENTHAL: And to clarify, in our piece of paper the discussion point says, quote, AThere was concern on how the developers handled right and left hip/knee replacement since there is limited ability to distinguish between right and left.@

Well, right and left is not the relevant question. It is unilateral or bilateral. Is that what the TAP meant by right and left? Or am I missing something?

DR. WEINSTEIN: Well, what we meant by right and left is you=ve got two knees, right? And so, are you doing one or two?

CO-CHAIR ROSENTHAL: Yes, okay, it=s the one or two that is the issue --

DR. WEINSTEIN: Yes.

CO-CHAIR ROSENTHAL: -- not whether they did the right one or the left one?

DR. WEINSTEIN: Correct.

CO-CHAIR ROSENTHAL: I=m just

clarifying.

DR. WEINSTEIN: Correct. Correct.

CO-CHAIR ROSENTHAL: Okay.

DR. WEINSTEIN: But if you follow cohorts of patients, and if they did two and you don=t know which one then got revised or readmitted for some other reason because of right or left, you can=t attribute it the same way. So, it is complicated.

CO-CHAIR ROSENTHAL: I was just going to comment that, if you look at administrative claims data for this, sometimes they don=t code whether it is the left or right. So, you don=t know if you are using administrative claims data.

And then, when you have two knees done, did they do the same knee again or did they do the other knee? And you don=t know for sure. So, there is some ambiguity in terms of the way the data flows in.

DR. WEINSTEIN: Yes, it is just that we should all be aware of this; that=s

all.

DR. PETER: I wonder if the ETG folks could talk about how the grouping function handles that issue?

MR. LYNN: Well, I can say that our experience has been that the bilaterality modifier is used more predictably than left and right. And we may already exclude cases that are bilateral. I would have to go back and check the detail on that. But we certainly could. It would be, I think, more predictable to exclude bilateral cases.

DR. WEINSTEIN: Yes, but you don=t know, if the modifier is there, if it is not there, you still don=t know sometimes.

MR. LYNN: Yes, that=s definitely true. But I think because the bilateral ones are compensated differently, I think --

DR. WEINSTEIN: That=s true. That=s true.

MR. LYNN: -- that they are more likely to, especially since it increases the

compensation, they are more likely to identify it; whereas, left and right doesn=t affect compensation.

CO-CHAIR ROSENTHAL: Yes, to me, it is the issue of bilaterality that you want to know. I would be surprised if people wouldn=t code for that because the payment changes. And I am still not quite sure that the right/left question is all that important in this thing, but --

DR. LEE: We run very few buy-one/get-one-free sales on the hospital side.

(Laughter.)

CO-CHAIR ROSENTHAL: But I am assuming, Jim, that this was the basis, then, though, for the TAP vote, which was, if I am looking at it correctly, zero high, 3 medium, and 4 low? Would that be fair to say?

DR. WEINSTEIN: Yes. Yes, I mean that is part of it, I think. So, the last part of the writeup there, lack of clarity on the procedure definitions, handling of

comorbidities and the weighting of the multiple comorbidities was the other issue. So, we definitely think this is important; i.e., the unanimous vote. But there were some problems with the methodology here on the measurement around reliability and how it was done. You know, it is fixable, but that=s what we said.

CO-CHAIR ROSENTHAL: Okay. Is there any further discussion on reliability?

(No response.)

Hearing none, then I think we will call the question. Would you guys mind reading again the TAP scores? And then, we will do our vote.

MS. YANAGIHARA: Is it possible to have Carlos= assessment.

CO-CHAIR ROSENTHAL: Oh, yes, I=m sorry. We should do that.

MS. YANAGIHARA: It will be helpful.

CO-CHAIR ROSENTHAL: And then, we

will do the second part. Thank you.

So, Carlos, would you mind giving your portion on reliability? Carlos? On mute?

(No response.)

He may have dropped off.

DR. BARNETT: The question I don=t think we mentioned here is the one thing that it said -- am I reading the right one, about the dementia? Is that right?

CO-CHAIR ROSENTHAL: I think it was on the other one, Paul, but go head.

DR. WEINSTEIN: That was on the hip fractures.

DR. BARNETT: That was on the other one. Sorry. Sorry.

CO-CHAIR ROSENTHAL: Dementia is not as much in the under-65 here. And I think, wasn=t the idea that Carlos would sort of give us the one, and his response was largely comparable across each of these measures that we are going to be considering

today? So, I think that is why he is not available for each individual one. You thought he would be on?

MS. WILBON: So, for 1609, for 2a1, whether or not the specifications were well-defined and precise, we had 3 moderate and 4 low. And for 2a2, which is on the reliability testing, we had 2 high and 5 moderate. Overall reliability, we had 2 high and 4 moderate.

CO-CHAIR ROSENTHAL: Okay. So, those are the TAP scores, and we are voting overall reliability, and a 1 is high; 2, moderate; 3, low, and 4, insufficient.

(Whereupon, a vote was taken.)

MS. FANTA: Okay. And for everyone on the phone, overall reliability, high, moderate, low, or insufficient.

I know Jeptha had to walk away for a minute. So, I don=t think he is on the phone right now.

But, Doris?

DR. PETER: Moderate.

MS. FANTA: Okay. Jim?

DR. WEINSTEIN: Moderate.

MS. FANTA: Okay. And Ethan?

DR. HALM: Moderate.

MS. FANTA: Okay. So, the final results are 2 high and 14 moderate.

CO-CHAIR ROSENTHAL: Near unanimity.

Okay. So, now let=s move to the next portion about scientific acceptability which will be the validity questions.

And so, Jim, would you give us the TAP view on validity?

DR. WEINSTEIN: Yes. We had some issues here. I want to make sure I=m covering the right ones.

But, as our comments state, the six months prior we thought might have been too long to incorporate in this group. And then, the question is, are we looking at system level or single provider, which goes

back to one of the questions you raised on the last one. At any organization level, does somebody do enough of these? We know there is tremendous variation in rates of these procedures by provider. Most joints that are done, people do less than 10 a year, or most of the people who do joints do less than 10 a year, which is kind of amazing. So, there were some problems there.

And I=ll stop there.

DR. PETER: Hi. This is Doris.

I just have a question. Since there is something like 70 percent of the costs were attributable to the hospital, if it is a provider-level measure, then I guess I was wondering what the variability is in the hospital rates because I wasn=t sure what the provider would do if the hospital is contributing to so much of the overall cost.

DR. WEINSTEIN: Doris, this is Jim.

I may be wrong, but I thought they

didn=t really specify their cost measures. They sort of just used the standardized price and cost. And maybe that was hospital, but I am not sure I remember that well.

DR. PETER: But even if it was standardized, it is still a percentage of the total cost. I guess I was trying to understand what the purpose of the measure was at the clinician level. I almost feel like for the physician it should be a rate level rather than a utilization level.

CO-CHAIR ROSENTHAL: I think we need to clarify this. We had a discussion earlier, and we probably should have clarified it on the previous measure. But all these Ingenix measures are total cost, just dollars.

DR. PETER: Right.

CO-CHAIR ROSENTHAL: So, these are not standardized priced. These would not take into consideration price differences from one hospital to another, one provider group to another. It is just the dollars.

DR. PETER: Right.

CO-CHAIR ROSENTHAL: Okay? And apparently, that was not the case when the TAP discussed this. So, you all would perhaps have been unclear on that point, but we spent 45 minutes or so at the beginning of this meeting this morning talking about that.

DR. PETER: Oh, my apologies. I missed the discussion this morning.

CO-CHAIR ROSENTHAL: Yes. Sorry.

Jim, would you comment? The one vote that I am looking at that was particularly skewed negative had to do with risk adjustment. Could you just elaborate on that a little bit?

DR. WEINSTEIN: On the reliability part or?

CO-CHAIR ROSENTHAL: Yes, under validity. No, it is 2b4 is risk adjustment.

DR. WEINSTEIN: 2b4?

CO-CHAIR ROSENTHAL: It says, the notes here are, AThere was a lack of clarity

on severity-level assignments and how they related to the risk adjustment model. The TAP agreed that not all the comorbidities provided in the submission seem appropriate for the population in the measure.@

DR. WEINSTEIN: Yes.

CO-CHAIR ROSENTHAL: Does that ring a bell?

DR. WEINSTEIN: Not as well as it should, I guess. But I don=t remember that.

CO-CHAIR ROSENTHAL: I have the benefit of the piece of paper.

DR. WEINSTEIN: I am just guessing, you know, severity is a hard thing. I don=t know whether you use radiographs for severity. I don=t know how that was done. I can=t remember that. I=m sorry.

DR. SINNOTT: This is Patsi.

And I was just looking. They used the DRG to define severity. So, depending on the DRG rating or categorization at the discharge, I think, that determines the,

quote, Aseverity@ of the case. But there are also issues about what are the other comorbidities that might be influencing outcome.

CO-CHAIR ROSENTHAL: All right. David, do you have a comment? And then, we will ask Ingenix to clarify. Or, David, and then Barbara, and then we will ask Ingenix to clarify.

DR. REDFEARN: This measure is unique in the sense that they don=t use the built-in risk adjustment that comes in the ETGs. They use MSDRGs, but it is not specified very well. That kind of ambiguity I think is what the TAP was responding to.

CO-CHAIR ROSENTHAL: Barbara?

DR. RUDOLPH: Yes, I just want to clarify on the numbers of knee replacements physicians do. If you go to the Massachusetts government site, about the lowest is 19 per year, and it goes up to 230. So, I don=t think there is really as big an issue with

this in regard to small numbers for physicians.

DR. WEINSTEIN: Yes, in the Medicare data, it varies a lot more than that.

CO-CHAIR ROSENTHAL: Ingenix, would you just comment on the risk-adjusting methodology?

MR. LYNN: Yes, I would. My colleague David Redfearn said it exactly right, that we don=t use our comorbidities, et cetera, for severity. We use the MSDRG for the admission.

CO-CHAIR ROSENTHAL: Was there some reason for that selection, for that choice?

MR. LYNN: Well, as someone else pointed out, these cases don=t have as much variability as the condition cases on the cases that involve a major anchor procedure like knee or hip replacement. And we felt like the severity risk adjustment was sufficient.

CO-CHAIR ROSENTHAL: ***So, based on the data, let=s just clarify; there=s a couple of puzzled expressions. So, if I am hearing you correctly, when you look at the overall populations that undergo these procedures, you are saying there is not a lot of variation and there is not a lot of variation that you see in the underlying comorbidities. Hence, the methodology required to, say, Aadequately@, in quotes, risk adjust is much less than you would need if you were looking at something like coronary artery disease or diabetes or one of the other conditions. Am I paraphrasing it correctly?***

MR. LYNN: ***I think that is exactly right.***

CO-CHAIR ROSENTHAL: ***Okay.***

MR. LYNN: ***If you look at the unadjusted distributions of, say, coronary artery disease versus an episode around a knee replacement, the coefficient of variance is much lower for the ones around knee***

replacement.

CO-CHAIR ROSENTHAL: Okay. Lisa?

MS. GRABERT: What if you are not paid on a DRG? What risk-adjustment methodology do you use for that? There=s a lot of people who are paid on APRDRGs or at a per-diem rate.

MR. LYNN: Our example showed how this could be done with MSDRG, but I think that the methodology says you are using a DRG measure. So, I think our methodology is written so that you could use MSDRG or you could use APRDRG.

CO-CHAIR ROSENTHAL: But, again, you end up with total cost. So, it doesn=t really matter. The measure is cost. So, Lisa, does it matter how it was paid, again, because you=re not measuring the underlying utilization.

MS. GRABERT: Right.

MR. LYNN: Well, I think she is talking about the severity adjustment method.

MS. GRABERT: ***Yes, for the severity adjustment it is. Because when you bill a DRG, you have nine --***

CO-CHAIR ROSENTHAL: ***Your point was around the risk adjustment, not the validity of the underlying -- are you comparing apples to apples once you have counted up the dollars? That was your point, yes, okay.***

DR. REDFEARN: ***And, of course, you don=t have to pay using DRGs. You pay on per diem. You can always run the MSDRG grouper on the same data to pull the risk adjustment out.***

MR. LYNN: ***Right.***

CO-CHAIR ROSENTHAL: ***Well, you can, but does the measure specify that?***

MR. LYNN: ***The measure specifies that you use a DRG, whether it is MSDRG or APRDRG or some other grouper, to help with severity adjustment.***

CO-CHAIR ROSENTHAL: ***All right. So, again, it seems to me, it is how the thing***

is specified is how it is supposed to be used, not how you could use it or it can be varied, or some customer might decide to customize it. What we are voting on, I think, is how it is specified on the pieces of paper in front of us.

Lisa, do you have another point on that?

MS. GRABERT: Yes, I would like to believe that, when you are paid on a per-diem basis, that those claims easily run through a DRG grouper, but the fact of the matter is they don=t. And you are going to get all kinds of errors that bounce back. So, I don=t know that that is a proper method for risk-adjusting non-DRG-based claims data.

CO-CHAIR ROSENTHAL: Is it fair to say that people who submit claims that are paid on per diems, there may be higher coding errors? Is that what you are suggesting? And then, when somebody has to translate it at the other end, you reiterate the coding errors as

you try to retranslate it back? That=s the point? Okay.

DR. REDFEARN: I can only comment for our data. We don=t see that at all. In California, we pay largely per diem and we routinely run MSDRGs and APRDRGs on the data, and we don=t see that problem. But that is our own particular situation.

CO-CHAIR ROSENTHAL: Any other discussion on the validity questions, either questions for the TAP, questions for Ingenix, discussion among the group?

(No response.)

Hearing none, then I would suggest one of us would again now clarify what the TAP votes were on the five subsections and then their overall vote on this. Then, we will take our own vote.

MS. WILBON: All right. So, for 1609, for the validity subcriteria, for 2b1, whether or not the specifications are consistent with the cost-of-resources problem,

2 high, 4 moderate, and 1 low. For validity testing, we had 1 high, 4 moderate, and 2 low. For exclusions, we had 2 moderate, 4 low, and 1 insufficient. For risk adjustment, 6 low and 1 insufficient. And for 2b5, the identification of statistically-significant and meaningful differences, we had 3 moderate, 2 low, and 1 insufficient. And then, the overall validity was 1 moderate and 5 low.

CO-CHAIR ROSENTHAL: Okay. So, our vote will not be on the subsections; it will be on overall validity. And again, the scoring is for us 1, high; 2, moderate; 3, low, and 4, insufficient.

So, with that, is everybody prepared to do their clickers?

And, Sarah, are you ready for us to go? Yes.

(Whereupon, a vote was taken.)

MS. FANTA: Okay, and for everyone on the phone, again, it is overall validity, high, moderate, low, or insufficient.

Jeptha, are you back?

DR. CURTIS: Yes, but I came through a little bit late.

MS. FANTA: Okay, no problem.

DR. CURTIS: So, I would like to abstain.

MS. FANTA: Doris?

DR. PETER: Moderate.

MS. FANTA: Okay. Jim?

DR. WEINSTEIN: Low.

MS. FANTA: Sorry?

DR. WEINSTEIN: Low.

MS. FANTA: Oh, low, okay.

Ethan?

DR. HALM: Low.

MS. FANTA: Okay. So, we have 1 high, 9 moderate, and 6 low.

CO-CHAIR ROSENTHAL: All right. So, now we need to vote on overall scientific acceptability. Am I correct? Help me, folks.

So, is there any further discussion about any aspects of scientific

acceptability, which then captures all the elements and all the gestalt around scientific acceptability?

(No response.)

And on this, it is 1 is yes and 2 is no.

So, if there is no further discussion, Sarah, are you ready?

(Whereupon, a vote was taken.)

DR. PETER: Are you all still there?

(Laughter.)

CO-CHAIR ROSENTHAL: Yes. One more time, everybody. One more time. Somebody is not -- yes, don=t point at Sarah; point at the end of the laptop out here.

Did we get it? We=re missing one person. Let=s do it again. One, yes; 2, no. We=re revoting. One, yes; 2, no.

Kurtis, reach out and really just reach around there one time at the end of the table because that is the most likely -- yes,

not to pick on anybody; it is probably mine.

Yes, there are 13 of us. All right, we failed twice. Nine, 10, 11, 12. All right, we are going to have to do a show of hands.

Okay. All the yes votes, please raise your hand. This will narrow it down.

(Show of hands.)

All right, noes?

(Show of hands.)

Wait. Let=s do it again. We can=t even do the hand votes. We=re missing a no. So, one of the four of us has got a faulty clicker. Okay.

DR. PETER: Because make people separate across the room, the yeses on one side and the noes on the other.

(Laughter.)

CO-CHAIR ROSENTHAL: All right. Now let=s get the phone votes.

MS. FANTA: And then, for everyone on the phone, yes, scientific acceptability,

yes or no.

Jeptha, I don=t know if you were able to listen. Do you want to vote on this?

DR. CURTIS: No, I will abstain from this.

MS. FANTA: Okay. Doris?

DR. PETER: Yes.

MS. FANTA: Okay. Jim? Jim, are you still there?

DR. WEINSTEIN: Yes. Sorry.

MS. FANTA: That=s okay. Yes?

DR. WEINSTEIN: Yes.

MS. FANTA: Okay. Ethan?

DR. HALM: No.

MS. FANTA: So, it looks like we have 11 yes and 5 no.

CO-CHAIR ROSENTHAL: All right. Now let=s move, then, to the usability question.

So, Jim, do you want to give us the TAP version of usability?

DR. WEINSTEIN: Yes. I think it

says in here, but we had a hard time following some of this formulaically and the hierarchy of the model. So, we think this is important and probably usable, but it is pretty complicated.

CO-CHAIR ROSENTHAL: Is it possible you could explain for us the difficulty around the complication?

DR. WEINSTEIN: Yes. Well, you know, the rankings, they are confusing. In some cases, the lowest number is the strongest association; in some cases, the highest number is the strongest association. This assumes coding is consistent between facilities. It isn=t always. And as I said before, it doesn=t always address or it doesn=t address specific resource utilization within a procedure or an E&M visit, things like that.

CO-CHAIR ROSENTHAL: There were a couple of puzzled looks in the room when you said the lowest and highest didn=t correlate. Would you mind explaining that?

DR. WEINSTEIN: ***I think even up above in the scoring analysis they talked about this winsoring thing. To me, winsoring means you sort of disregard or discard equal values on both sides. And they sort of just took the low outliers and excluded them and not the high outliers, those kinds of things. I wondered about the usability because of the methods and whether they were valid in that sense.***

CO-CHAIR ROSENTHAL: ***Okay. Questions, then, from the group? There have to be some because there=s lots of puzzled looks.***

DR. NEEDLEMAN: ***Yes, I read the comments from the TAP and I think the Ingenix response on that. The Ingenix response made sense to me. They thought the really low -- and we are talking very, very low -- charges represented miscodings of the primary diagnosis. And the winsoring at the upper end is just they have standard practice for***

bringing the extraordinarily high charges down to their cutoff level. Both of those seem to be reasonable judgments in how to deal with the data.

MR. AMIN: Can I offer a piece of clarification, Tom, because I know that there is some confusion here? And, Jim, please correct me if I am wrong.

Some of the TAP concern here was around, they had a large discussion around the strength of association of how individual claims would be assigned to various concurrent episodes. The response from Ingenix was around the tiebreaker logic that is used in their model. And the TAP expressed they were uncomfortable with the lack of the clarity that was provided on the tiebreaker logic and the strength of associations.

I don=t know that that helps clarify or further complicates, but I offer that.

CO-CHAIR ROSENTHAL: I didn=t

follow. What is the issue with the tiebreaker methodology? Can you explain that, Taroon?

MR. AMIN: The tiebreaker, I don=t know that I can explain, but what I can explain is there was a lack of clarity around how the tiebreaker logic works and, also, because it was explained that there is a level of strength of associations that were provided in the tables, and these strengths of associations were not clear to the TAP in the evaluation of how individual claims would be assigned to concurrent episodes. Is that clear?

CO-CHAIR ROSENTHAL: To concurrent episodes? How would you have --

MR. AMIN: As part of the risk-adjustment model.

DR. SINNOTT: Well, this is Patsi.

You could have two concurrent episodes not necessarily the same thing. So, your patient who has a total hip replacement done gets pneumonia in the hospital.

CO-CHAIR ROSENTHAL: I got it. And which episode do you attribute it to then?

DR. SINNOTT: Is that a different episode or is that part of the hip fracture episode?

CO-CHAIR ROSENTHAL: All right. And I have a question, based, again, on what is in the paper that is in front of us, which says, AThere was concern that this episode is not being currently used or reported as a standalone measure. As such, the developer was unable to provide any data on its current use as an individual measure.@

Does that mean this has not been tested in any real-life situation?

DR. WEINSTEIN: That=s what we understood.

MR. LYNN: This is Tom from Ingenix.

We have used it in real-life situations. We have not used it as only a measure for a hip replacement or only a

measure for a knee replacement. It is used by our customers as a composite measure --

CO-CHAIR ROSENTHAL: And what is the composite measure?

MR. LYNN: -- in other procedures as well.

CO-CHAIR ROSENTHAL: What is the composite measure, Tom?

MR. LYNN: The composite measure would be that you would look at it alongside of other knee procedures that were done, other orthopedic procedures that were done by that group or that physician.

CO-CHAIR ROSENTHAL: So, in other words, you have in use around your customers total orthopedic care or total orthopedic procedures?

MR. LYNN: Right.

CO-CHAIR ROSENTHAL: But not hip and knee replacement specifically?

MR. LYNN: We don=t have as many folks, looking only at hip replacement or only

at knee replacement.

DR. SINNOTT: Can you drill down to it, though, if you want to see it?

MR. LYNN: Oh, yes, you can drill down to it.

DR. SINNOTT: Yes. Okay.

MR. LYNN: Yes, so we have experience. That=s my point really. Thank you. Just to sort of solidify it, my point is we do have experience using this measure. Just most of our customers use it as a composite with other measures.

CO-CHAIR ROSENTHAL: Okay.

DR. SINNOTT: To measure the performance of a physician who is classified as an orthopedic surgeon, for example, or a group?

MR. LYNN: For example.

DR. SINNOTT: Or a group. So, this is Patsi again.

So, my personal comments about this were that the measure is used in various

forms for public and private reporting, but we don=t know whether they have been useful to users, and we don=t know, because the clinical logic about classifying and episodes -- you know, ultimately, you want these instruments, these scoring functions to be useful to physicians specifically, so that they can understand how their practice is varying from their peers. And if the clinical logic is not transparent -- and maybe David could speak to this and how they have used it -- if it is not transparent, then the physicians can always say, AWell, my patients are sicker.@

And we did not get enough information about the clinical logic that went into the classification to be able to infer that it would be useful to either administrators or providers.

CO-CHAIR ROSENTHAL: So, Patsi, you are saying, if I am hearing you correctly, that one of the tests, in your mind, for usability is that it has actually been used

and, as it were, validated against the real world, where those being measured are telling us back that they accept the judgment of the measure, as it were?

DR. SINNOTT: So, yes.

CO-CHAIR ROSENTHAL: This has not been put through that test?

DR. SINNOTT: Well, it is not so much that it has not been put through that test because I think, for example, that WellPoint uses it for various functions within their management of physician performance and incentive bonuses and things of that nature. What we didn=t get in the reporting was information about how it is used, you know.

So, we don=t know if it is meaningful. We don=t know if, for example, the physicians have said, AWell, this is a great tool. We like this, and we will go ahead with it,@ or AWe=ll put up a big uproar about it and say we don=t think this is valid. Therefore, we are going to your using it.@

***CO-CHAIR ROSENTHAL:** Okay, yes, I follow you. That certainly seems to be at least one criteria for discerning whether something is usable or not.*

Lisa?

***MS. GRABERT:** I have a statement and a question for the developer. I thought early on, as a Committee, we decided that we weren=t going to review composite measures because this is a new body of work and it is a difficult area, which I think that this measure does serve as a composite measure.*

Aside from that, my question for the developer is, what is your client=s reason for combining these two procedures? Is it a small numbers issue? Why don=t they look at these procedures individually?

Because when I ran this data on the Medicare program, we always separated out these two procedures with ETGs. We didn=t combine them in a composite.

***MR. LYNN:** Yes, I think it is to*

get adequate numbers is part of it, but, also, part of it is, you know, at the end of the day, many of our customers want a score that represents a provider=s practice or a group practice or a system=s practice. And that is why you would use a composite to do that.

And then, again, like you pointed out, there is an ability to drill down to see what procedures are drilling the composite score one way or the other.

CO-CHAIR ROSENTHAL: But could I clarify? Tom, I thought I heard you say that this measure, meaning hip and knee replacements, in your typical customers are rolled up into multiple other orthopedic procedures which are the composite to which you were referring, not this measure being a composite of hip and knee replacement?

MR. LYNN: Right, that=s true. We would roll it up further than just hip and knee replacement. You know, thinking off the top of my head -- and I don=t know every

single customer -- it turns out, for this particular rule, I think we do have at least one customer that looks at like major joint, and actually probably rolls up just these two rules. But, for the most part, our customers roll up more than just whatever rule we are discussing.

CO-CHAIR ROSENTHAL: And is that because even hip and knee replacements don=t typically generate enough material in your customer base to provide a meaningful comparison of cost between one orthopedic surgeon and another?

MR. LYNN: I think in some of the cases of some of the rules that is probably true. For the case of this rule, it is probably less true. It is more about trying to get to a single measurement for a system or a group or a provider.

DR. SINNOTT: Well, can I suggest, also, that you wouldn=t want to be evaluating whether a physician or a surgeon was in or out

of your panel based on a single procedure? You would want to see their experience across the procedures that take up most of their time and cost you the most money.

MR. LYNN: That=s right.

CO-CHAIR ROSENTHAL: All right. Any further questions or discussions on the usability question?

(No response.)

If not, I am going to suggest that we vote. The voting, as I understand it, on this is high, moderate, low, and insufficient.

If you will give us the TAP scores on this, then we will do the vote.

MS. WILBON: Sure. For the TAP 3a, which was the measure performance results are publicly reported, there was 5 moderate and 2 low. For 3b, measurement results are meaningful and useful for public reporting or performance improvement, that was 4 moderate and 3 low. And 3c, the data results can be decomposed or deconstructed for transparency

and understanding, 3 moderate and 4 low.

CO-CHAIR ROSENTHAL: All right. So, we will apply the same scoring system. One is high; 2, moderate; 3, low; 4, insufficient.

And, Sarah, if you are ready?

(Whereupon, a vote was taken.)

MS. FANTA: And for those of you on the phone, usability, high, moderate, low, or insufficient?

Jeptha?

DR. CURTIS: Low.

MS. FANTA: Doris?

DR. PETER: Moderate.

MS. FANTA: Jim?

DR. WEINSTEIN: Moderate.

MS. FANTA: And Ethan?

DR. HALM: Moderate.

MS. FANTA: So, we have zero high, 12 moderate, 4 low, and 1 insufficient.

CO-CHAIR ROSENTHAL: All right. So, on the home stretch on this measure, now

we have feasibility.

Jim?

DR. WEINSTEIN: Yes. To go back to my sheets here, you know, I think part of the discussion that just occurred was some of the confusion I had myself. I don=t know about the rest of my colleagues, but I was thinking of trying to get to a measure that was useful for an individual doc, too.

And I understood the discussion, but I think the discussion we had around feasibility that it states there was that data elements only routinely generated in the care process. I=m not sure that that happened here, and I need to look back at the actual documents to see what I was referring to, unless somebody wants to help me out with memory.

MS. WILBON: Well, Jim, this is Ashlie.

4a and 4b, we didn=t spend a lot of time on, seeing as how all these measures

use admin data.

DR. WEINSTEIN: Yes.

MS. WILBON: So, 4a, which asks whether or not the data elements are routinely generated, admin data, most people would agree, is routinely generated. And then, for 4b, whether or not the data elements are available electronically, also, most admin data is available, most or all admin data is available electronically.

But if you want to focus on 4c and 4d, 4c being about the susceptibility to inaccuracies and unintended consequences, and then, 4d, whether or not a data collection strategy can be implemented and about any barriers to use there may be.

DR. WEINSTEIN: Yes. Well, I think it says that in the statement there. The issue here to me, again, this issue of preferences, one of the things -- and it is one of my own biases -- that rates of procedures may look good on paper, but we

don=t know that patients who are well-informed actually want it or that they had other options. And it gets to this issue of people who are sort of conservative people might look like outliers. You know, they are only treating different kind of patients.

I am not sure that I capture this in this group or in this modeling because, again, I am confused now that this doesn=t actually get down to the individual doc on a total knee or a total hip replacement. Some people do just that.

CO-CHAIR ROSENTHAL: But I think it does get down to individual docs, not for knees versus hips, but for total between hips and knees it would attribute the cost down to the individual doctor level.

DR. WEINSTEIN: Yes, but what about the doc, as I just said, who doesn=t do a lot of surgery, who just sees a lot -- an orthopedic surgeon who is very conservative? I mean he just would be seen as a very low-

cost person.

CO-CHAIR ROSENTHAL: No, but this is not a capitated measure. This isn=t cost of care against a group. Then it would be relevant. But this is, if you do a hip replacement, what does it cost?

DR. WEINSTEIN: Yes, yes.

CO-CHAIR ROSENTHAL: So, it is not taking into consideration at all appropriateness, but it doesn=t purport to. If it were a capitated measure, then the issue about appropriateness, your point is still well-made. You could have a situation where a surgeon doesn=t do very many and, consequently, is very conservative, but when he does one, is expensive.

DR. WEINSTEIN: Right, right.

CO-CHAIR ROSENTHAL: That could be an unintended consequence because the guy who is expensive on a per-case basis is really saving a group or a health plan or something a ton of money because he or she is, in fact,

incredibly conservative about who they elect to operate on.

DR. WEINSTEIN: Yes, yes.

CO-CHAIR ROSENTHAL: But that is going to be inherent in any procedurally-based costing consideration. And I am assuming some people are going to still find it useful to know the per-cost number.

So, were there any other feasibility questions? Because the feasibility largely pertains to the point of, can you get the information you need without a lot of hullabaloo? And this one doesn=t seem to be terribly different than any of the others that rely on administrative data, other than issues that would relate to its reliability or its usability, but in terms of feasibility, this, to me, seems pretty straightforward.

Jack?

DR. NEEDLEMAN: I just want to get some clarification of what the concern over

cost is. If it is a matter of operating versus not operating, then I think you are right, this measure does not capture that, and that=s fine. It doesn=t purport to do that.

So, what I need to understand from the clinicians in the room is whether, if you are being conservative, so you are operating on folks that are in more pain or more disability in some sense, is it going to be a more expensive treatment than if you are operating on folks that are in from that extreme level? Or is it the same cost once you have decided to operate?

DR. WEINSTEIN: I think some people would argue that -- maybe I didn=t say it very well -- some people would argue that; it is that my patients are sicker. But, in this case, they have a worse disease, and so they may be more complicated to fix and the surgery may take longer, and the utilization of resources may be different.

But I would be curious what other

people think.

***CO-CHAIR ROSENTHAL:** Yes, but that all may be true, and I don=t know, but if it were, the time to have considered it was under validity and under was it accurate --*

***DR. WEINSTEIN:** Yes, yes.*

***CO-CHAIR ROSENTHAL:** -- not under whether it is feasible.*

***DR. WEINSTEIN:** Yes, I understand.*

***CO-CHAIR ROSENTHAL:** And is the risk-adjusting adequate to take that all into consideration without creating a skewed or inaccurate rank ordering of people?*

***DR. WEINSTEIN:** I get you, and I=m not sure --*

***CO-CHAIR ROSENTHAL:** So, if it was an important question, we should have asked it 10 minutes ago.*

***DR. WEINSTEIN:** Yes.*

***CO-CHAIR ROSENTHAL:** Kurtis?*

***DR. ELWARD:** Yes, hopefully, I can clarify. Speaking somewhat objectively, as a*

primary care doctor who sees my patients back after the surgeons get done with them, I think that, in general, it will work itself out. There are some people who are in a tremendous amount of discomfort and they sail through the operation and do fine, and other people who have been getting by and they just happen to have a different pain threshold. So, I think it will, overall, average out.

DR. BARNETT: Just speaking to the feasibility issue, it is kind of an interesting approach. In order to do this Ingenix process, you have to run the episode grouper on all your data because you have exclude the care that is, for instance, the pneumonia episode that occurs concurrently with a hip replacement operation. So, that is their way of dealing with case mix, in essence, is by building the episodes and excluding the care that is not relevant to the specific replacement.

So, the feasibility issue is you

have got to work with all the data and episode group all the data. And so, the alternative would be to look at some larger costs of care beyond just the episode and then do a case mix control, which also requires looking at all the data to see whether they had concurrent pneumonia. But you would perhaps include that cost in the alternative.

So, really, in terms of feasibility, it is how comfortable you are with the idea that you have got to run all the claims data through the episode grouper in order to get at just this issue. And so, it may be an equivalent amount of data that you have to look at, and then it is a question of how much you trust the episode grouper versus some other measure of risk adjustment like HCCs or what other people have used for other measures, what NCQA is doing, for instance, with some of the measures that they have proposed to us.

So, that, to my mind, is the

difficult thing about feasibility, is you have got to get this whole product running for all of the episodes that it can create in order to just answer this one question.

CO-CHAIR ROSENTHAL: Helen, I might ask your counsel at this point. We have considered at one point in time the cost-of-the-product question. And quite honestly, I can=t remember quite exactly how it played out, but there may be people involved in this discussion that haven=t been involved previously. And it probably is worth some statement around that. So, whether that is you or Ashlie at this point, but that would be appropriate to do at this point.

So, Ashlie?

MS. WILBON: Right. So, in 4d, the whole data collection strategy and barriers to use are identified and include looking at whether or not there are any fees associated to use, whether or not the data is accessible, and so forth. So, within that

subcriteria is where we had asked before that you guys review the fee structure that Ingenix submitted. So, that would also be a consideration for this subcriteria for all the Ingenix measures.

What I was going to suggest is if maybe we would bring up or just kind of recall for you guys how you voted on other Ingenix measures on feasibility, because in a lot of ways this criteria should be consistent across all the Ingenix measures. I think a lot of the issues are probably the same.

So, to kind of speak toward consistency, or I=m not sure how you want to handle this.

CO-CHAIR ROSENTHAL: Well, if you recall, we didn=t vote on the feasibility ones at the last live meeting. We did them on the phone call because we didn=t have the fee structure. So, maybe you could both remind us of the fee structure and remind us how we voted after the phone conversation?

MS. WILBON: Sure.

CO-CHAIR ROSENTHAL: To the extent that internal consistency is a virtue of a committee, we can at least look at that.

MS. WILBON: Sure.

CO-CHAIR ROSENTHAL: So, why don=t you tell us both the fee structure and how we voted?

MS. WILBON: Sure. I can bring up the fee structure. And actually, the results that we showed earlier this morning, when we talked about the costing structure, in there was actually a feasibility vote. So, we can share that. Just give me a second to pull that up.

CO-CHAIR ROSENTHAL: And in the meantime, I will just reiterate Paul=s point was, and again, to the degree that it is relevant, the issue about feasibility is you can=t run this measure in a vacuum. You virtually have to run all of your data through the grouper in order to parse any of them out.

Barbara?

DR. RUDOLPH: That being said, it is unlikely that someone would just do this one measure. So, if we are going to look at cost, I think most of the people who are going to use this already have APRDRGs. They probably already have the MSDRG stuff set up. And it is not an enormous deal to push the data through it.

CO-CHAIR ROSENTHAL: Yes, I guess, in my own mind, and I get the point of view of, if it is valid that anybody can use it, is one person being able to use it sufficient for us to endorse it, or are we endorsing this as a kind of national measurement that we would expect to be widely implementable? And I don=t think we have ever really resolved that question here. And I think there=s even perhaps differences of opinion.

Would we have ever endorsed a quality measure around, say, pressure ulcer rates with the notion that, well, only three

places in the country are actually capable of measuring pressure ulcers, but we are going to endorse it anyway because it is otherwise a valid measure, but only three places really can use it? I don=t know whether we would have for most of the quality measures if there are any that work like that, but maybe I=m wrong.

DR. RUDOLPH: I think there are a number of registry measures that only those who have the registry data can use.

CO-CHAIR ROSENTHAL: Ashlie, how close are we to --

MS. WILBON: Pretty close.

CO-CHAIR ROSENTHAL: Pretty close. And then, we will vote, and then we will have lunch. Well, I guess we have to vote overall acceptability, and then we will have lunch.

But we are making very good progress. I mean we are way ahead of schedule here.

MR. BOWHAN: Can I ask a question

about the measures that we are talking about? We seem to be talking about them just in actual prices, but in the Ingenix system it seems like they have that cost per episode, but they also have an index. And is that also included?

CO-CHAIR ROSENTHAL: No. I mean, again, I think that --

MR. BOWHAN: That=s not part of this?

CO-CHAIR ROSENTHAL: -- what we heard, again, if I am understanding your question correctly and the discussion we had this morning, this is only the cost; this is not the index.

MR. BOWHAN: Well, I mean the cost is part of doing an index. So, anyway, that is what I wanted to be clear on, whether or not what we are talking about -- because when you get to the comparison part, that is where you have it, is in the index. You don=t have it in just the pure cost measure.

CO-CHAIR ROSENTHAL: Maybe we should clarify. When you say Aindex@, what do you mean?

MR. BOWHAN: Well, they calculate your score for an individual provider. And then, what they do is they take your peers in that area and they average costs for it. So, you get an expected.

And then, to the discussion about, gee, if someone is being more conservative and they are only seeing more severe patients, when you looked at the index, you would be comparing apples to apples. And if this measure includes both the dollars per episode as well as the index, then you can get to where you want to go.

CO-CHAIR ROSENTHAL: Okay. All right. Can anybody clarify that? Ingenix, do you want to clarify that? I think it does include an index, correct, in the way that Jack just described?

MR. LYNN: Yes, he described that

very well.

CO-CHAIR ROSENTHAL: Okay. So, that=s the answer. The answer is yes.

DR. NEEDLEMAN: But the index is based on the risk adjuster, correct? So that, when you are making the adjustments for cost per episode, you are looking at the different risk-adjustment categories. And if I understand the risk adjustment on this one, it is based on with or without comorbidities and complications. And if that is correct, none of those relate to the severity of the illness, the severity of the underlying condition, because that is not included in those codes. It is simply is there some other comorbidity or complication in the care that is bumping up the cost of the treatment because they had pneumonia or they had diabetes or they had dementia or some other thing that bumps you into the higher DRG category.

So, what I heard was, yes, we are

giving people the index based upon expected, but the expected is based upon the risk-adjustment model, which includes other conditions, but doesn=t include variations in severity of illness within the hip or knee.

CO-CHAIR ROSENTHAL: I believe that is all accurate and was perfectly appropriate for the conversation when we discussed scientific validity.

(Laughter.)

We are now discussing feasibility, for which none of this is relevant. Pardon me.

But we are all killing time here, anyway.

(Laughter.)

We are just waiting for Ashlie to find out what our previous feasibility vote was.

So, what I believe you have got on the screen, although it is a total blur to me --

MS. WILBON: Yes, we can read it.

CO-CHAIR ROSENTHAL: -- but this is the dollars for the various sized groups to remind us of the cost part, and then you are going to give us the validity --

(Pause.)

MS. WILBON: Obviously, we have it on the screen, but it is very hard to see. So, I am going to just read it aloud.

So, for the ETG, again, this is a recollection of how they price their product for the ETG, depending on the size of the provider. So, they divide it up by small, medium, and large. It ranges from 70K to 110, and this is for a three-year term and does not include installation and annual fee for a three-year term.

Oh, I=m sorry, that was for MDs, for physician groups. And then, for a plan, they also divided it up into small, medium, large, and then, by commercial and government. Then, the range for commercial is 90 to 135,

and then, for government it is 65 to 100. And that is just for the ETG. So, again, they have ETG, ERG, ETGPG, but for this particular measure, only the ETG pricing would apply.

CO-CHAIR ROSENTHAL: Okay. And then, our previous votes on feasibility? Yes, feasibility.

MR. LYNN: I don=t need to make a statement again, just for interest -- but this is Tom Lynn -- and this particular rule requires ETG and TAG and I think ETG, to add TAG is to add like 30 percent to the cost.

MS. WILBON: Okay. Thank you for that clarification.

Yes, plus installation. Okay.

So, for feasibility, there were four Ingenix measures that you guys voted on. And so, it is actually pretty consistent, the way you voted on feasibility. You generally had about two to three high, mostly concentrated in the medium and low, moderate and low ratings.

So, for 1591, on feasibility, we had 2 high, 8 moderate, 7 low, and 1 insufficient.

For 1594, which was a CAD measure, we had 3 high, 8 moderate, 6 low, and 1 insufficient.

For the diabetes, we had 2 high, 8 moderate, 8 low.

For the non-condition-specific, we had 3 high, 8 moderate, 6 low, and 1 insufficient.

So, actually very consistent.

CO-CHAIR ROSENTHAL: So, let=s see how we do now.

MS. WILBON: Right.

CO-CHAIR ROSENTHAL: Now that we know all of this information, I think we are prepared to vote, and all that very good conversation.

So, it is 1, high; 2, moderate; 3, low, and 4, insufficient, and this is on feasibility.

(Whereupon, a vote was taken.)

MS. FANTA: And then, for those of you on the phone, on feasibility, either high, moderate, low, or insufficient.

Jeptha?

(No response.)

Jeptha, are you there?

(No response.)

Doris?

DR. PETER: Yes, moderate.

MS. FANTA: Okay. Jim?

DR. WEINSTEIN: Moderate.

MS. FANTA: Okay. And Ethan?

DR. HALM: Moderate.

MS. FANTA: Okay. Thanks.

So, we have 1 high, 8 moderate, and 7 low.

CO-CHAIR ROSENTHAL: All right. Either we=re wonderfully consistent or a foolish consistency is a hobgoblin of little minds. I guess only time will tell.

(Laughter.)

All right. So, we have gone through all of the submeasure components, then, of this measure. And I think now it is time to vote on overall acceptability. And so, this is recommendation for or against endorsement. So, it is either yes, no, or abstain. So, 1 is yes; 2 is no; 3, abstain.

Point of order, Lisa?

MS. GRABERT: I actually wanted to make a comment before we called a vote on this. I was looking through the documentation again. And sorry, I have to go back to the composite issue again.

Because the cost on average for a hip episode is about $2,000 less than a knee episode because they are two separate, distinct episodes that have been combined in a composite measure. So, if you happen to have a physician that has got more hip or more knee, there is not really a fair comparison when you use a composite measure for these two different episodes.

So, my question is --

MR. LYNN: No, no, no, that=s not true. I hate to interrupt you, but they have different expected values when you create the ratio.

MS. GRABERT: So, do you weight differently between the two episodes when you put them in a composite? Is that how you address it?

MR. LYNN: Yes. So, it is actually a little bit more complicated. And just to sort of simplify it, to be quick, if you have a hip, you have an average cost of $5,000; if you have a knee, it is an average cost of $6,000 across the peer group. Then, your cost goes in the numerator and the average cost for the peer group goes in the denominator for the calculation of the ratio. So, that is what allows you to compare things or include in one ratio things that are different.

CO-CHAIR ROSENTHAL: So, it

accounts for the fact that a particular surgeon or a particular group might have a different percentage of hips or knees?

MR. LYNN: Right, and uses the same extrapolation of what I just said to take into account that one doc may have a bunch of knees with comorbidities and complications on the DRG and another one may not.

CO-CHAIR ROSENTHAL: Okay.

MS. GRABERT: Tom, can you refer me to the page in the specification document where that is spelled out?

MR. LYNN: I can, but it will take me some time.

CO-CHAIR ROSENTHAL: All right. I think we are ready to then call the vote on overall recommendation for endorsement. So, again, just to clarify, now we are recommending or not recommending endorsement of the whole measure. And this will be 1 is yes, 2 is no, and 3 is abstain.

So, Sarah, are you ready?

(Whereupon, a vote was taken.)

MS. FANTA: Okay. And for those of you on the phone, overall recommendation, yes or no.

Jeptha?

(No response.)

Doris?

DR. PETER: Yes.

MS. FANTA: Jim?

DR. WEINSTEIN: Yes.

MS. FANTA: Ethan?

DR. CURTIS: Yes.

MS. FANTA: Thanks.

Okay. So, we have 9 yes and 7 no.

CO-CHAIR ROSENTHAL: All right. I think we are finished with this measure, and it is time for lunch.

MS. WILBON: So, for those on the phone, we are going to break for 30 minutes, and we should be back at about 1:15.

CO-CHAIR ROSENTHAL: 1:15.

MS. WILBON: So, we will continue

with the pulmonary measures after lunch.

DR. SINNOTT: This is Patsi.

Did you do the low back pain?

MS. WILBON: Before we break for lunch, we do need to have public comment.

So, Tom, if you are still there on the phone, if there is anyone on the participant line who would like to make a comment, now is the time to do so.

THE OPERATOR: And all lines are open.

(No response.)

MS. WILBON: Is there anyone there who would like to make a comment?

(No response.)

Okay. Great. Thank you.

So, we are now officially breaking for lunch.

DR. BARNETT: So, that is the last of the bone/joint measures? We are not taking up the back pain, to answer Patsi=s question?

MS. WILBON: That=s correct.

Because they were included in the summary, we still included it in the summary because it actually happened, but ABMS withdrew after that meeting. So, just the two Ingenix, and then we will move on to the pulmonary measures after lunch.

DR. SINNOTT: Okay. Thank you.

MS. WILBON: Thank you.

MR. LYNN: What time are we reconvening?

MS. WILBON: About 1:15.

MR. LYNN: Okay. Thank you.

MS. WILBON: Thank you, Tom.

MR. LYNN: You probably already said that. I apologize.

MS. WILBON: No, It=s fine. Thank you.

(Whereupon, the foregoing matter went off the record at 12:43 p.m. and resumed at 1:26 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

1:26 p.m.

CO-CHAIR ROSENTHAL: All right, I think we will get started.

We are on Item 1611, the ETG-based pneumonia cost-of-care measure from Ingenix.

So, Tom, if you are still on from Ingenix and would want to give us a very quick overall of this? And then, we will move to the TAP discussion.

MR. LYNN: Yes, I would just point out that this is a disease or condition rule. So, therefore, just using the ETG technology with a severity adjustment of the ERG or PEG. And that is treated as an acute disease. So, it has that moving window like the hip fracture did.

I think that=s it.

CO-CHAIR ROSENTHAL: Okay. Kurt, are you in charge of the TAP on these? So, I am going to ask if you would sort of give us a quick overview, and then we will get to each

of the segments. Is there any sort of overview that you would want to give about this measure?

DR. ELWARD: One of the interesting things about the approach is just this episode-based concept, which I think is intrinsically interesting. One of the challenges we had as a TAP is to look at the measure carefully to see if the wide range of clinical presentations of pneumonia, you know, the different sources and the treatments, could be captured as well. That is, of course, a significant challenge.

I think, as you can see in some of the different -- we thought everything was important in all the measures that we will be presenting this afternoon. You will see a fair amount of variability across some of the measures. And particularly, we probably need to talk a little bit about the usability and how that impacts availability, following all the discussion this morning.

Ingenix was very good about following up on the questions we had, and we can go through those.

But that is about all I have to say right now.

CO-CHAIR ROSENTHAL: Okay. I think that is a good start.

So, this is ETG-based pneumonia resource use measures. So, the first question would be importance.

Is there any discussion from anybody on the Committee that they want to have about importance?

(No response.)

If not, let=s go through the formality of 1 is yes; 2 is no.

And, Sarah, are you ready?

(Whereupon, a vote was taken.)

All right, try again.

We didn=t lose anybody from lunch, did we?

Most of this is to just test to

see that the system is working. Otherwise, I am going to call -- did you get it? Okay.

MS. FANTA: And for those of you on the phone, for importance for 1611, yes or no.

Jeptha?

DR. CURTIS: Yes.

MS. FANTA: Jim? Jim Weinstein?

(No response.)

Okay.

CO-CHAIR ROSENTHAL: It sounds like we lost him.

MS. FANTA: Doris? Doris, are you there?

(No response.)

And Ethan?

DR. HALM: Yes.

MS. FANTA: Thanks.

All right. So, we have 14 yeses and 1 no.

CO-CHAIR ROSENTHAL: Okay. So, let=s now move to scientific acceptability,

and we will start with 2a and 2b, reliability.

Kurt?

DR. ELWARD: Yes, overall, we felt somewhat uncomfortable with the lack of transparency in the risk-adjustment specifications. The severity weights, particularly for the elderly, were unclear. And there were these clean periods where you count the utilization for a while, and then, finally, there is decrement in the utilization. That seems to open things up for a new episode.

CO-CHAIR ROSENTHAL: Can you explain that a little bit, what you mean by clean period?

DR. ELWARD: Well, perhaps the person from Ingenix can help me out.

MR. LYNN: Yes, that sounds fair.

So, a clean period with acute diseases is basically a time period, I believe pneumonia it is 60 days. And basically, if you have an interaction between a clinician

and a patient around pneumonia, then that clock, that 60-day clock, starts. If you have another encounter within those 60 days, it is not just 60 days; the clock restarts every time a clinician and a patient get together and the issue is pneumonia. And so, the episode continues until there is 60 days where there is no pneumonia activity for that member.

DR. ELWARD: I do think that Ingenix did a good job of explaining how that works. It is an intrinsically-complicated process, though.

It is important in that you don=t want to keep accruing charges for something that may have nothing to do with regard to pneumonia. So, the advantage of that -- correct me if I=m wrong -- is that just because you have pneumonia, and happen to have a bunch of other things going on, you don=t continually get those charges, that resource accumulation.

CO-CHAIR ROSENTHAL: So, this is the stopping rule for the episode.

DR. ELWARD: Exactly. That is the easiest way to do it, yes, it is the stopping point.

CO-CHAIR ROSENTHAL: But was the TAP satisfied that the stopping rule made sense in light of the way pneumonia works in relationship to, say, other intercurrent diseases, et cetera?

DR. ELWARD: Yes. I think our sense is that we wanted more clarification from them on particularly some separation between community-acquired and healthcare-acquired pneumonia, since they were very different clinical situations. I think we were still requesting that they give us a little bit more detail in how that would work.

CO-CHAIR ROSENTHAL: So, Tom from Ingenix, can you comment on the difference between community-acquired and healthcare-acquired pneumonias, and how that is accounted

for in the model?

MR. LYNN: Yes. So, the model uses diagnosis information to do its grouping, uses some diagnostic information from procedure codes, but doesn=t use the procedure codes themselves to try to categorize disease. The risk there is you don=t want sort of utilization to drive it, to be one of the markers that you use to determine high cost. So, that is why we didn=t see how to distinguish those two things without using utilization as a marker, which we were trying to avoid.

CO-CHAIR ROSENTHAL: But I guess the question I am hearing posed is that they are potentially two different diseases. And therefore, you would have to a priori distinguish them in order for ultimate comparisons to be valid.

It is the same question or a similar question to the one around hips and knees. If, in fact, one is vastly more

expensive than the other, and you don=t have a way of identifying so that you could weight them, what if one set of providers has more of one kind of pneumonia than the other?

Kurt, am I phrasing the question correctly?

DR. ELWARD: Right. Exactly.

MR. LYNN: Yes, I guess our answer to that is that, to the extent that is reflected in diagnostic information, it is taken into account in this in the severity adjustment. But if it is not, then it is not.

And we understand the risk of saying, well, this happened to you in the hospital. Then you are sort of using utilization to determine high cost.

CO-CHAIR ROSENTHAL: You know, I get it becomes a circular argument, and you don=t want to do that.

MR. LYNN: Right. So, that is what we were up against, basically.

CO-CHAIR ROSENTHAL: Right. I get

that.

Kurt, for the non-clinicians among us, how significant an issue is this going to be in terms of having a homogenous or of this being now viewed as an inhomogenous population?

DR. ELWARD: Yes, correct me, you know, Taroon and Ashlie can correct me if I=m wrong. I think it was still a significant issue. And it depends on how the health system is able to splice their own data. If they know which is which and you separate the two, then it will fine. If it is a group measure, I think one of the issues is that people who get hospital-acquired pneumonia are usually intrinsically sicker than the people who get community-acquired pneumonia. So, I think unless there is a way of separating those out, that it is going to be a continuing problem.

CO-CHAIR ROSENTHAL: Okay. And the risk-adjusting component, because you also

have some concerns -- well, I guess we will get to that in the next part of the thing, about the risk adjustment, as to whether that is sufficient to pick that up or that you would still really want to know a priori which kind of pneumonia you actually were dealing with. And since this is all coded data, it is not coded, is that --

DR. ELWARD: Right.

CO-CHAIR ROSENTHAL: Right.

Yes, Barbara.

DR. RUDOLPH: If it is hospital-acquired pneumonia, wouldn=t that be reflected in like present on admission versus community-acquired?

CO-CHAIR ROSENTHAL: It might be if the coding is really accurate. It might be if the coding was really accurate.

And so, to the extent that you could make the same argument about any of the things we are dealing with; I mean it all depends on the coding being accurate. The

question is, is the present on admission around pneumonia a more significantly badly-coded kind of thing or not? And I don=t know the answer to that. I think you make a good point.

DR. REDFEARN: This is David Redfearn.

I think Tom was referring to the fact that that would be considered utilization, the fact that you are admitted to the hospital and they excluded that because they didn=t want utilization to come into the definition. Maybe Tom can correct me if I got that wrong.

MR. LYNN: Yes, David, that is the point I am making about --

CO-CHAIR ROSENTHAL: Yes, but I think Barbara=s point was that, if, in fact, you got admitted and the code was community-acquired pneumonia, you would, in fact, know that this one was community-acquired and not hospital-acquired.

MR. LYNN: Right.

CO-CHAIR ROSENTHAL: But it is not universally true because, again, I don=t know how accurately that is identified as present on admission. Otherwise, all you would get is a discharge --

MR. LYNN: Well, and we are not using present on admission, either.

CO-CHAIR ROSENTHAL: Okay. So, there=s the answer to that one.

MR. LYNN: We could. We have the same concerns you do about present on admission. I think that it is, from my experience even with Medicare, it is pretty dicey. And I think commercially it is not even used ubiquitously.

DR. ELWARD: I think overall the Committee was convinced that, given that hospitals are pretty good about coding those things because they do have significant relationship to reimbursement and safety measures, that as long as the coding by the

hospital was correct, that the measures would be good.

CO-CHAIR ROSENTHAL: My experience, though, is like what Tom just described. The present-on-admission codes are very, very badly used because they are not at the moment related as much to reimbursement, with a few exceptions for Medicare, and in the commercial world they are not terribly applicable. And so, I don=t think most hospital coding for present on admission is done particularly well.

DR. ELWARD: It would have to be based on discharge diagnosis.

CO-CHAIR ROSENTHAL: Right.

DR. ELWARD: Yes.

CO-CHAIR ROSENTHAL: And the discharge diagnosis is going to be pneumonia.

DR. ELWARD: Well, they should be able to -- I think there are different codes for different types of pneumonia.

CO-CHAIR ROSENTHAL: Okay.

DR. ELWARD: So, you classified them --

MR. LYNN: Yes, just to clarify that, there are different diagnosis codes for different pneumonias, and we do take those into the account into the building of our severity. But I don=t think there is like a diagnosis code for hospital-acquired. It is just you can tell from the organism pretty well.

CO-CHAIR ROSENTHAL: In your TAP, you talk about lack of transparency with the risk-adjusting specifications, but I think if we can postpone that until the validity discussion where the risk adjustment is called out?

Are there other questions then of the TAP or around the reliability questions specifically? Anybody from the Committee?

Yes, Jack.

DR. NEEDLEMAN: Just I would like a reaction from the folks who were on the TAP.

There was concern about transparency, and Ingenix responded with some comments. As you read their response, are you comforted? How comforted are you?

DR. ELWARD: Is Janet on the line?

DR. MAURER: Yes, I just came on the line. So, I am not quite sure what we are doing here.

DR. ELWARD: Yes, what we are talking about is, in terms of transparency, how comfortable were we in the end? I think we were comfortable enough that, and you can see the scores, some of the exclusions are very good. Some of the replicability seemed very good, and validity in terms of the evidence being consistent with intent was good.

When you got into risk adjustment, there is much more concern about how we could open up the box and see what is in there.

MR. AMIN: Kurt, maybe I can add some additional detail there.

DR. ELWARD: Please.

MR. AMIN: I think a lot of the discussion here was also derived from the statistical review of the measure around whether there was sufficient level of detail around the specific techniques and the multi-variate regression about how specific variables were included and excluded and the calibration and goodness-of-fit details. So, the R-squared value specifically was asked for by the TAP.

And there was a response provided. Now the level of that response to answer these questions is up to interpretation. But those were the concerns that were addressed or brought up by the TAP during the discussion.

DR. ELWARD: Thank you. Thank you.

CO-CHAIR ROSENTHAL: Well, if the R-squared was asked for, what was the answer? So, we don=t know?

DR. ELWARD: No, I don=t think we

got one on that.

DR. MAURER: I don=t see any answer on these.

CO-CHAIR ROSENTHAL: Okay. All right.

MS. ZIELINSKI: Hi. This is Cheri from Ingenix.

We provided the R-squares in our response to the followups. Ashlie, did you not receive those?

MS. WILBON: Are those in the Word documents you sent?

MS. ZIELINSKI: Correct.

MS. WILBON: Yes, we did receive those, and we passed those on. So, I would have to look in detail. I am not really sure -- so, I think they are looking at the one for 1611, and it doesn=t appear to be in there.

CO-CHAIR ROSENTHAL: Well, if I could make a suggestion on behalf of the group, the risk-adjusting methodology by our

kind of guidance falls under the reliability, although it is obviously -- I mean under validity, although it is obviously also relevant to the reliability question. But maybe I could suggest that we discuss it in detail, and maybe in that length of time somebody can discern whether or not we actually got the figures from Ingenix, and that we could take the reliability question on its own without the statistical validity.

People okay with that? All right.

CO-CHAIR STEINWALD: Our input from Carlos is limited to what he talked to us about before lunch, but it is understood to pertain to all of the Ingenix measures?

MS. WILBON: Yes. So, when I talked to him about it yesterday, he said that the methodology and approach they used for reliability/validity testing for all their measures is consistent across all of them. So, there was very rarely anything that was very different about any one of the measures.

CO-CHAIR ROSENTHAL: ***I guess the question I have around that is, and it isn=t clear to me, intuitively, something like hip and knee replacement seems to have a tight, would have a kind of tighter degree of fit because the start and stop rules ought to be more obvious in relationship to the way clinical care actually happens, and that the things like congestive heart failure, which the group I guess we did not endorse this morning, and one like this one, might have less clear-cut starting and stopping rules, might have more intercurrent kinds of things, and therefore, might not be as tight as something like a procedurally-oriented thing.***

But Carlos did not make any differentiation himself around that particular point?

MR. AMIN: ***While all that would be accurate, Tom, I think the question that was raised by the group was just what is the R-squared, not necessarily comparing the***

R-squared against procedural fits. So, I think that level of detail was requested. So, that is what I think the challenge is.

MS. ZIELINSKI: Hi. This is Cheri with Ingenix again.

I am looking at the followup items that were requested from us for pneumonia. And the specific R-squared scores weren=t asked for. There were four followup questions, and none of them were asking for the specific R-squared.

So, we can produce those. I don=t think we would have a problem with producing those. But I just wanted to be clear that for pneumonia this is not one of the four items that was asked for us to deliver.

MS. WILBON: It might not have been specifically listed for pneumonia, but it was asked for for all the measures. But that is a fair statement. It might not have been that specific one, but throughout the conversation it was requested. So,

understood.

CO-CHAIR ROSENTHAL: Okay. So, I would suggest, again, we will consider the risk-adjusting thing under the next heading. And I would suggest let=s read the scores from the TAP on reliability, and then we will vote on that section.

So, who is going to relate this to us? Scores? Scores, so we can vote.

MR. AMIN: Okay.

DR. BURSTIN: I just want to let people know that these slides were emailed to them. So, if they want to pull it up, if you have email, you could see it at your own little desk, if that is easier to read.

MS. WILBON: And we moved the screens closer. So, hopefully, people can see them a little bit better.

CO-CHAIR ROSENTHAL: All right. It is better, but just help us.

MS. WILBON: Yes.

CO-CHAIR ROSENTHAL: It is pretty

apparent now what=s what.

So, 2a?

MR. AMIN: 2a1, well-defined and precise specifications, 3 high, 4 moderate. Reliability testing, 6 high and 1 moderate.

CO-CHAIR ROSENTHAL: Okay, and then overall?

MR. AMIN: Overall, 3 high and 3 moderate.

CO-CHAIR ROSENTHAL: Okay. And, Kurt, if you don=t mind, could I just ask one more question? Then, we will vote on the thing.

Interestingly, in the measures we talked about before lunch the discussion from the TAP seemed to me to be fairly benign, whereas, the scores were not so good. And on these, the discussion felt a bit more negative, and yet, the scores seem pretty high.

Are there inter-rater reliability issues or am I missing -- does my question

make sense?

***DR. ELWARD:** No, I think the fact that we were still concerned initially about how you open this up, I think for the most part Ingenix gave good answers in how we dealt with it. So that we thought the reliability by people using it, if they knew how to use it, was high and moderate, and overall, the validity was moderate.*

***CO-CHAIR ROSENTHAL:** Okay.*

***DR. ELWARD:** And actually, there were very few lows.*

I think one of the reasons behind the discrepancy is that there is a challenge for the individual user who is trained and knows these data, and knows the measures, they can probably do really well. The challenge for us was to say across plans, if you start comparing different plans, you can get into some challenges as far as do they really understand what --

***CO-CHAIR ROSENTHAL:** I guess that*

would get into the usability part --

DR. ELWARD: Yes.

CO-CHAIR ROSENTHAL: -- which we discussed again before lunch.

DR. ELWARD: Usability was a big deal.

CO-CHAIR ROSENTHAL: I guess what you are also saying is that all of these grouper-oriented methodologies produce kind of challenges because they are not all in the public domain and they have not all been analyzed by an army of statisticians and readily understandable. So, consequently, one group could look at it and see it somewhat differently than another.

DR. ELWARD: Yes. And again, Janet, maybe you can help me out on this.

DR. MAURER: Yes. So, before lunch, it looks to me like you talked about procedures, right, hip and knee, and so on?

CO-CHAIR ROSENTHAL: Yes.

DR. MAURER: And now, this is,

like someone mentioned earlier, a very different situation where you have acute illnesses that might not have as good of start and stop dates, and so on. And I think there is a little more discomfort in working with these medical illnesses than with the procedure-oriented issues, and especially in the setting where you are trying to assign cost using a specific episode of a specific illness.

So, I think it is understandable that there would be a little more concern about how that is done.

CO-CHAIR ROSENTHAL: Yes, it just wasn=t reflected in the scores.

DR. MAURER: Well, I mean, you did have a different team doing the other ones, though.

CO-CHAIR ROSENTHAL: Yes, that=s okay. That=s all right. It wasn=t reflected in the scores in the same way that the earlier ones were. But that=s not a big issue.

Okay. So, I am ready to call the question on overall reliability, 1, high; 2, moderate; 3, low; 4, insufficient.

Sarah, are you ready?

(Whereupon, a vote was taken.)

DR. ELWARD: I wasn=t sure mine was working. So, I did the other one. So, take two off. Okay, I won=t do it anymore.

(Laughter.)

Now that I know both work, I=m in good shape.

As they say, vote early and often.

(Laughter.)

So, you can put 10 for moderate.

MS. FANTA: Those of you on the phone, overall reliability, high, moderate, low, or insufficient.

Let=s see. Jeptha?

DR. CURTIS: Moderate.

MS. FANTA: Okay. And Ethan?

DR. HALM: Low.

MS. FANTA: Low. Okay.

So, it looks like we have 2 high, 11 moderate, and 2 low.

DR. PETER: And hi. This is Doris. You can add me, too. I=m moderate.

MS. FANTA: Oh, sorry. I didn=t know you were back.

DR. PETER: No, it=s okay.

MS. FANTA: Sorry, that was moderate?

DR. PETER: Yes, it was.

MS. FANTA: Okay.

CO-CHAIR ROSENTHAL: All right. Great. Let=s move on, then, to --

MS. FANTA: So, 12 moderate.

CO-CHAIR ROSENTHAL: Thank you.

MS. FANTA: Sure.

CO-CHAIR ROSENTHAL: Let=s move on to validity.

Kurt?

DR. ELWARD: Yes. The overall validity was moderate. There was a little bit more discomfort in some of the measures. The

risk-adjustment methodology is inherently -- you know, that=s software that they run. And so, it is not readily transparent. They did seem to have a good command of how they were doing risk adjustment, and I think Carlos felt like they were doing a very good job.

But, as mentioned in the TAP discussion, we still were concerned that certain types of pneumonia couldn=t be separated out.

So, the overall validity as a general measure for pneumonia was felt to be moderate, but we still had concerns about the fact that it was hard to separate different types.

CO-CHAIR ROSENTHAL: Okay. Questions? Paul?

DR. BARNETT: So, just to understand, a person could be immune-suppressed or have heart failure and develop pneumonia. And so, does the pneumonia episode, the cost of this pneumonia is

associated to their immune-suppressed disease? Say they have HIV disease or heart failure. Or is this a new pneumonia episode?

DR. ELWARD: No, what Ingenix -- and maybe the Ingenix people can fill in -- but the understanding that was given to us was that that risk-adjustment methodology does, in fact, include those things, which is one thing that is inherently helpful about it. And although it is a complex process, those comorbidities are factored in.

CO-CHAIR ROSENTHAL: So, can we clarify that from Ingenix?

DR. BARNETT: Is it a pneumonia episode or is it an HIV episode if it has pneumonia as a comorbidity --

MR. LYNN: It is a pneumonia episode with a comorbidity of HIV.

CO-CHAIR ROSENTHAL: And so, can we just clarify the other obvious things that would create a more significant pneumonia, like other forms of immune suppression or

transplantation or a variety of things? Are they in the risk-adjusting methods?

MR. LYNN: Right, they are in the risk-adjusting methods as comorbidities. yes.

DR. WEINSTEIN: So, let me ask, does the method distinguish a patient with HIV who has pneumocystis pneumonia versus a patient with HIV who has pneumococcal pneumonia or an opportunistic infection from a run-of-the-mill community-acquired infection?

MR. LYNN: Right. So, there is a condition status which is an internal marker to pneumonia, and there is one for pneumocystis and one for pneumococcal pneumonia.

DR. ELWARD: I would say that was one strength of the Ingenix data, is that they go into quite a bit of detail accounting for different types of pneumonias, which, on the one hand, may not be as applicable for community-acquired pneumonia; for conditions such as HIV, it might.

CO-CHAIR ROSENTHAL: Okay. Jack?

MR. BOWHAN: I hope I am bringing this question up at the right time this time.

(Laughter.)

CO-CHAIR ROSENTHAL: Go ahead. I=m sorry, I am not trying to be a stickler, but --

MR. BOWHAN: The index versus the resource use cost per episode, and maybe I can get this clarified, then, from Ingenix, if someone else around the table doesn=t know, so when they produce a number for the episode dollars, the dollars per episode, I don=t think any of that, the risk-adjustment factor or the severity plays into that number. It is only into the cost-of-care index where you are talking about severity and risk adjustment.

And is that a correct statement? I=ll ask the Ingenix people.

MR. LYNN: Yes, so what happens is that each of these markers contributes to a real number which represents the severity of

the episode, and where a 1 means the average episode across all pneumonias and a 1.2 means that the markers indicate a need for 20 percent increased utilization for this episode.

Those scores are then put in buckets, you know, just based on having a threshold. Below .8 is in severity level 1, and between .8 and 1.2 in severity level 2, et cetera.

And then, those buckets are used to create indexes across peer groups. So, how much did the average case across all the entities being evaluated cost for pneumonia in the different severity level groups?

Then that number is used as the expected value for an entity=s case of pneumonia, what their severity level is for that particular case of pneumonia. Their actual cost, of course, is put in the numerator, and the expected cost for that severity level across the peer group is put in

the denominator.

CO-CHAIR ROSENTHAL: Does that answer your --

MR. BOWHAN: So, just to clarify, the risk and severity adjustment only applies to the cost-of-care index, not to the resource use dollars per episode?

MR. LYNN: Oh, I=m sorry. No, we use it in all of those things.

MR. BOWHAN: So, if it cost a thousand dollars per episode, that number has been risk-adjusted and severity-adjusted?

MR. LYNN: No, the dollar amount is not severity-adjusted. The indexes are severity-adjusted.

CO-CHAIR ROSENTHAL: So, somebody could produce a ranking that had Jack=s cost as a provider of treating pneumonia of $2,000 and mine of $1,000, and those numbers could appear on a list without having been risk-adjusted?

MR. LYNN: Well, I mean, you could

do that, but I mean the measurement is the index.

CO-CHAIR ROSENTHAL: Okay. Well, I am just trying to clarify. I am not trying to argue. I am just trying to clarify.

MR. LYNN: And I=m sorry.

DR. REDFEARN: When you do the comparison, you would normally do the comparison within risk categories. So, if there are three levels of severity, you would say this doctor has AX@ number in this episode of pneumonia at severity level 1 and his average cost was this. And you would compare that average cost to the average for that episode and that risk level in his peers. And the same for level 2 or 3, or however many there were.

CO-CHAIR ROSENTHAL: How many levels of severity are there in the model? Four? Okay. Okay.

Other questions on overall validity?

Yes, Steve.

MR. PHILLIPS: Yes, I apologize if this in the materials. It is not popping out at me.

But I guess tying it back to the conversation this morning about hip fractures and the population and the proportion that is over 65, was that an issue here? I mean, do we have that breakdown?

DR. ELWARD: We did ask them about the difference in elderly particularly, and they did provide some response, which it appears that they have looked over what the difference would be and they can adjust by age. They actually didn=t find that that made a big difference in their model.

Am I correct on that?

MR. LYNN: This is Tom Lynn from Ingenix.

This is we are asking for approval in the commercial population. We did have in our data some folks that were over 65 where we

had all the information on them. It was much less than the commercial data. I mean we did develop separate markers for that age group, but I wouldn=t imagine we would have had a bunch of super-elderly patients like 75, 85 years old, or 85 years old.

DR. MAURER: This is Jan Maurer.

I think that, in general, across these measures it was the feeling of the Committee that there wasn=t probably adequate testing in the Medicare age patients.

MR. LYNN: Again, we are not asking for a recommendation for the Medicare. We are asking for commercial.

DR. ELWARD: Yes, they did, in the responses, they did identify a separate group of risk markers investigated, and this led to separate risk models based on elderly status for some conditions, for example, CHF and diabetes. But I don=t think, it sounds like they didn=t have enough data to really say that they could adjust this for the elderly in

a sufficient manner.

MR. LYNN: That=s correct.

CO-CHAIR ROSENTHAL: Okay. We=re sorting out the noise.

Jack, did you have a question?

DR. NEEDLEMAN: Yes. Is there enough homogeneity in this category of pneumonia that we can be looking at resource use across different kinds of pneumonias once the risk model is into account?

I am not being very clear here. Is it a single category that actually works or is there heterogeneity here that we should be worried, that I, as a non-clinician, should be worried about?

DR. ELWARD: I think -- and, Janet, you can correct me -- I think looking at hundreds of thousands of people, it would probably work. Overall, you get a picture of what resource use were, if you were looking at resource use and saying, where are your dollars going?

And they can identify out specific types of pneumonia, but I think we would like to see that better developed. And again, it hasn=t been tested that way.

CO-CHAIR ROSENTHAL: Well, but that does get to the question, and I have trouble with this because this issue of attribution falls down under usability in kind of our guidance on the thing. And yet, it cross-reacts, clearly, with the scientific acceptability --

DR. ELWARD: Exactly.

CO-CHAIR ROSENTHAL: -- and particularly validity.

Because I guess the attribution here is like the other attributions, which is it is specified down to the individual physician level, correct?

DR. ELWARD: Yes, I believe so.

DR. MAURER: It can be.

CO-CHAIR ROSENTHAL: So, is it accurate at the individual physician level,

given the various heterogeneities and the sophistication of the risk adjustment?

DR. MAURER: Well, the risk adjustment takes into account some of the situations where -- someone mentioned, you know, suppressed patients getting pneumonia. They would fall into the severity 4 level, as I understand it.

So, you have some risk adjustment that occurs that way. Are all hospitalized pneumonias homogenous? No, they are not. However, you know, community-acquired pneumonia that gets hospitalized is going to be a severe pneumonia. It is a little different from a hospital-acquired pneumonia. Does it differ in terms of the organism that is causing the pneumonia? Not so much. Maybe a little bit with Legion L or something like that.

But I think that their use of the severity level helps to distinguish immuno-suppressed-type opportunistic infections, say,

from those that might be just severe community-acquired pneumonias.

I don=t have a big problem with that. I think we need to see how it plays out in the real world when they are used.

CO-CHAIR ROSENTHAL: Tom?

DR. LEE: I mean, I don=t know whether Helen or NQF has any quantitative insight into this, but I have the impression that many hospitals that are performing well on a lot of quality measures look like they are doing badly on pneumonia quality measures. I mean I don=t have data, but it is the kind of thing that could be looked at, like for some kind of consistency thing.

Now one possibility, if that is true, is that maybe they are good on everything but bad on pneumonia. Another possibility which I think a lot of my colleagues suspect is that the pneumonia measures are problematic and more subject to coding issues. And if the quality measures

are skewed in that way, one could expect the resource measures to be skewed similarly.

So, that is why I have been sort of voting in a sort of skeptical way about these things in general. But that would be an interesting paper, actually -- (laughter) -- to see if the pneumonia measures are really running different compared to other quality measures at a hospital level. At a doctor level, I=ll bet you it is completely random.

***CO-CHAIR ROSENTHAL:** Well, and that is the problem. I am not sure the question really was answered that Jack posed and I added onto, which is, is this one going to be reliable down to an individual physician level, given the vagaries of the disease and the adequacy of the risk-adjusting? Whereas, they may be perfectly fine, as you are pointing out or suggesting, at a group level or a large level, but an individual physician --*

***DR. ELWARD:** I think at the*

individual physician on a lot of things, but particularly in this, yes, there would be big problems.

CO-CHAIR ROSENTHAL: But this is specified down to the individual physician level.

DR. ELWARD: And the reason it is, again, and not to defend them at all, but the reason it is is so that an individual health plan or a large group could go down and drill that down internally. But it would not be appropriate --

CO-CHAIR ROSENTHAL: Right, but not for public reporting or something like that.

DR. ELWARD: I think public reporting would be a huge problem.

DR. MAURER: Yes, this could be reported at the hospital level, too, though, could it not?

CO-CHAIR ROSENTHAL: Well, it could be, except it is specified at the

individual physician level. So, again, we can only approve it as it is specified. We don=t, I don=t think, get the opportunity to sort of revise it on the fly here.

MS. WILBON: You could ask them to change their level of analysis so that it would only be used at the higher level. So, it is basically like a checkbox that they check to say which levels of analysis it could be used.

DR. MAURER: Yes, one of the issues with reporting these at the physician level is that multiple physicians take care of that patient during a hospitalization. And this comes out in the NCQA measures, I think. So, that is one of the difficulties also of reporting at the physician level.

DR. REDFEARN: And what is the attribution rule here on this one? I don=t think we specified that. Does anybody know? Or, Ingenix, can you tell us?

MR. LYNN: Oh, I=m sorry.

CO-CHAIR ROSENTHAL: Yes, go ahead.

MR. LYNN: It is the same as the other ETG-based rules. It is based on activity. There either are contacts between a clinician and a patient or a total cost for a clinician and a patient. Either one of those methods can be used.

CO-CHAIR ROSENTHAL: But here, unlike the one we heard this morning around hip and knee replacement, where the attribution could only be to an orthopedic surgeon, I assume this one could be attributed to a primary care physician, a pulmonologist, a cardiologist.

DR. MAURER: An intensivist. I mean there could be many people who are taking care of this patient in the hospital.

MR. LYNN: Right.

CO-CHAIR ROSENTHAL: But it gets attributed, actually, though, to end up at the end of the day to one --

MR. LYNN: Right.

CO-CHAIR ROSENTHAL: -- which has the most --

MR. LYNN: Now there are threshold rules that are applied.

CO-CHAIR ROSENTHAL: Okay.

MR. LYNN: So that, you don=t assign a case to -- I think in this in our analysis we used 30 percent. We don=t assign a case to a provider, even if they are the highest, if they are not responsible for 30 percent of the visits or 30 percent of the cost, depending upon the method that you use.

And there was something else I wanted to say, but I can=t remember. That=s all right.

CO-CHAIR ROSENTHAL: All right. If it comes back to you --

MR. LYNN: It will probably come --

CO-CHAIR ROSENTHAL: -- just interrupt us.

All right. Anybody else have any other questions or comments or concerns that they want to ask, raise, or discuss?

Yes, Helen.

DR. BURSTIN: I don=t really think it is inconsistent. I mean there is certainly enough data to suggest that for some of these conditions we are seeing lots of different variability based on readmission mortality, for example. I haven=t seen anything specific for pneumonia. There have been a lot of pneumonia process measures that go to the clinician level already endorsed. So, that is pretty consistent.

I must admit, as a general internist, that doesn=t bother me. There is sort of one person usually who is the attending for a patient with pneumonia or somebody who has written that prescription. So, I am not seeing this terribly differently, just speaking out of turn as a clinician, but it is worth a paper.

(Laughter.)

CO-CHAIR ROSENTHAL: All right. So, can we get the TAP review scores here? And then, we will call the question.

MS. WILBON: Sure. So, for the validity subcriteria, you have 2b1, that the specifications are consistent with a resource use or cost problem. We had 4 high, 3 moderate --

MR. AMIN: Can I just clarify that, in 2b1, this would not reflect the change in the costing method that we discussed this morning. So, this would now be using actual cost and not offering the option of both.

CO-CHAIR ROSENTHAL: Okay. Thanks.

MS. WILBON: Validity testing, which is 2b2, we have 4 moderate and 2 low. For 2b3, which addresses exclusions, we had 2 high, 4 moderate, and 1 low. For the risk-adjustment subcriteria, 1 high, 3 moderate,

and 2 low. And then, for 2b5, which addresses the identification of statistically-significant and meaningful differences, we had 7 seven moderate.

CO-CHAIR ROSENTHAL: And did we get an overall --

MS. WILBON: Sorry. So, the overall validity was moderate, 7 moderate.

CO-CHAIR ROSENTHAL: Okay. All right. If there is no further discussion --

CO-CHAIR STEINWALD: There=s no further noise from the ceiling, either.

(Laughter.)

CO-CHAIR ROSENTHAL: It was beginning to sound like my dentist drill and having kind of the same impact.

So, now we are voting overall validity, and this is 1, high; 2, moderate; 3, low, and 4, insufficient.

Sarah, are you ready?

(Whereupon, a vote was taken.)

We=re missing one again. One of

us is a real miscreant.

MS. FANTA: And for everyone on the phone, we are voting on overall validity, voting high, moderate, low, or insufficient.

Jeptha?

DR. CURTIS: High.

MS. FANTA: Okay. Doris?

DR. PETER: Moderate.

MS. FANTA: Ethan?

DR. HALM: Moderate.

MS. FANTA: Okay. Thanks.

So, we have 1 high, 13 moderate, and 2 low.

CO-CHAIR ROSENTHAL: Did somebody on the phone vote high?

MS. FANTA: Yes. Yes, Jeptha. Yes, we have 1 high --

CO-CHAIR ROSENTHAL: Okay.

MS. FANTA: -- and then we have 13 moderate, and 2 low, and zero insufficient.

CO-CHAIR ROSENTHAL: Okay. So, now we are tasked to vote on overall

scientific acceptability. And this is 1, yes; 2, no.

Is there any further discussion before we do this?

(No response.)

Hearing non, Sarah?

(Whereupon, a vote was taken.)

MS. FANTA: And on the phone, voting on scientific acceptability, yes or no.

Jeptha?

DR. CURTIS: Yes.

MS. FANTA: Okay. Doris?

DR. PETER: Yes.

MS. FANTA: And Ethan?

DR. HALM: Yes.

MS. FANTA: Okay. So, we have 13 yes and 3 no.

CO-CHAIR ROSENTHAL: Okay. So, now we can move on to usability.

Kurt, the TAP on usability?

DR. ELWARD: Overall, the scores clustered around moderate.

Multiple organizations are currently using the measure. So, obviously, it is usable. And they have used it fairly consistently.

The challenge, not to rehash the different types of pneumonia that can be an issue, although it probably is more appropriate in the above-mentioned discussion, individual organizations could probably use this very well, but our major concern was that it would be difficult to use in a comparative setting across different large health systems.

And in some ways it depends on whether you are asking about usability in terms of can a large health system use it and estimate their cost and their utilization or whether you want to compare all the health plans in Chicago across each other.

CO-CHAIR ROSENTHAL: Would you mind elaborating a little on that just a little, if you could?

DR. ELWARD: Yes, I=ll try. I=ll

try to be clearer then.

CO-CHAIR ROSENTHAL: Okay.

DR. ELWARD: The thought was that multiple organizations currently use it, and many of them are finding it very usable in terms of their ability to look at their data. The measure would probably not be useful in a comparative setting.

CO-CHAIR ROSENTHAL: Comparing what to what?

DR. ELWARD: For example, if you were to compare two different organizations across --

CO-CHAIR ROSENTHAL: And what would make it not comparable, accurate in a comparison or usable in a comparison?

DR. ELWARD: Well, I thought was, and the reason they got moderate, is because if two groups who were still using the same, who were using Ingenix measures would probably find them comparable. The challenge would be it wasn=t clear how it would be used, say, if

somebody got a dataset on a bunch of different health centers that were not using Ingenix and just started comparing across health centers.

CO-CHAIR ROSENTHAL: Oh, so you are talking about literally the issue that we have raised on each one of these --

DR. ELWARD: Yes.

CO-CHAIR ROSENTHAL: -- Ingenix ones, that you literally have to use their product in order --

DR. ELWARD: Yes.

CO-CHAIR ROSENTHAL: Okay. Well, we have discussed that.

DR. ELWARD: Yes, right.

CO-CHAIR ROSENTHAL: We have a lot.

DR. ELWARD: But usability for individuals who have bought the software and use it, it seems to be usable.

CO-CHAIR ROSENTHAL: Okay. All right. And you are satisfied down to the individual physician level?

DR. ELWARD: I would say the previous discussion about keeping it larger would be better. I don=t know whether it would be -- obviously, people are using it and comparing individual physicians. It just takes the extra step of sorting out the individual variables.

I should say one thing. In taking this in context, we were trying to compare the other NCQA measures, which are very general, and which rightly suffer from not having any of the episode-based care. So, if you have pneumonia, you could be at risk for a lot of utilization that has nothing to do with pneumonia.

So, we were trying to look at this in the context of the very broad-brush approach that almost everybody else has used versus the attempt to be a little more defined that Ingenix is using. And we weren=t happy with much of it, but we were trying to put that in context.

CO-CHAIR ROSENTHAL: Okay. Other questions for Kurt or comments about usability?

DR. MAURER: I have one comment. This is Jan Maurer.

One of the issues that came up with respect to the usability across health plans was that standardized pricing was not used in the development of this. Although for any individual area, they do an observed-to-expected sort of expenditure use. And I don=t know that if you tried to compare that across regions where you could maybe use the observed-to-expected ratio okay, but certainly you couldn=t use just the cost because they would vary a lot.

CO-CHAIR ROSENTHAL: Okay. Yes, the thing we spent 45 minutes on this morning.

Okay. Any other questions or comments?

I=m sorry, please, Dolores.

MS. YANAGIHARA: So, this, again,

is in the commercial setting. So, I guess the question, again, is the numbers of pneumonia cases in the commercial setting, is that sufficient to get down to that level? It is kind of the same question as before, but --

CO-CHAIR ROSENTHAL: We never did get the R-squared, but I think we are just not going to have it. And I don=t know the answer.

Does anybody know the answer? Does anybody have an opinion about the answer? Opinion, if we can=t have facts, by God, we=ll have opinions.

(Laughter.)

Yes, please.

DR. RUDOLPH: I think with all the people who have asthma, who get bronchitis, and others, who are young, it seems there would be enough cases.

CO-CHAIR ROSENTHAL: I think there is a lot of pneumonia in a commercial population.

DR. MAURER: Yes, this goes up to age 64.

CO-CHAIR ROSENTHAL: Yes, I think there is a lot of pneumonia in that age group, yes.

I remain concerned about the attribution question down to the individual physician level, but, you know, it is hard to adjudicate --

DR. MAURER: That might be an issue, but certainly across a hospital, say, for example, you ought to get enough pneumonia.

CO-CHAIR ROSENTHAL: But it is hard to adjudicate that when you haven=t looked at the raw stuff. And we have been basing our decisions on this level of accuracy thus far. So, I don=t think we can avoid the question because we don=t have every last fact on it.

All right, 1, high; 2, moderate; 3, low; 4, insufficient.

Sarah?

(Whereupon, a vote was taken.)

MS. FANTA: And for those of you on the phone, we are voting on usability, high, moderate, low, or insufficient.

Jeptha?

DR. CURTIS: Insufficient.

MS. FANTA: I=m sorry, what was that?

DR. CURTIS: Insufficient.

MS. FANTA: Okay.

CO-CHAIR ROSENTHAL: Did anybody hear him?

MS. FANTA: Yes.

CO-CHAIR ROSENTHAL: You bet. Okay.

MS. FANTA: Yes. Doris?

DR. PETER: Moderate.

MS. FANTA: Thanks.

Ethan?

DR. HALM: Moderate.

MS. FANTA: Okay. So, we have 3

high, 11 moderate, 1 low, and 1 insufficient.

CO-CHAIR ROSENTHAL: All right. I am going to suggest that we not spend much time on feasibility. We have discussed the feasibility issue around the Ingenix thing to death.

Kurt, unless you have something really substantial to add to that, or anybody else has a burning issue around feasibility?

(No response.)

I do think we are obligated to vote on it. Are we going to consider that the vote is --

MS. WILBON: We can carry that vote forward for the remaining --

CO-CHAIR ROSENTHAL: Are people comfortable with carrying the previous feasibility votes forward and not per se voting again on feasibility? Okay.

MS. WILBON: Is everyone okay --

CO-CHAIR ROSENTHAL: Is everybody okay with that, as a point of order?

Okay. So, then, we are left now to vote overall acceptability and recommendation or not for endorsement, and the vote here is 1, yes; 2, no, and 3, abstain.

And so, Sarah?

(Whereupon, a vote was taken.)

MS. FANTA: Okay, and on the phone, overall endorsement, yes or no.

Jeptha?

DR. CURTIS: Yes.

MS. FANTA: Thanks.

Doris?

DR. PETER: Yes.

MS. FANTA: Okay. And Ethan?

DR. HALM: Yes, reluctantly.

(Laughter.)

MS. FANTA: Okay. Thanks.

So, we have 12 yes and 4 no.

CO-CHAIR ROSENTHAL: All right. That concludes the discussion on 1611.

Now we will move to 1605. Or do we want to break? We don=t need a break.

I=m just being asked in the background about consistency, and we can talk about that at the break or we can sleep on it a little bit. Because why pneumonia and not congestive heart failure? But I would say let=s postpone asking that question. We either have to be perfectly consistent or we can tolerate a modicum of inconsistency. I am not sure what the justification is between pneumonia -- but let=s ponder on that for a moment. But rather than trying to address it cold, move through and deal with the asthma measure.

Then, we will take a quick break. Then, we should be able to get finished.

So, it is 1605.

Kurt, are you ready?

All right. Well, let=s take 30 seconds and everybody get ready.

And, Ingenix, while he is getting ready, do you want to give us the 30-second version on the asthma measure?

MR. LYNN: Absolutely.

Asthma is based on the ETG technology. It has severity adjustments that are similar to the other rules. And it is the timing of chronic disease was divided into year-long episodes.

CO-CHAIR ROSENTHAL: All right. Kurt, are you ready?

DR. ELWARD: It was clearly felt that it was very important, and I think all of us endorsed that.

CO-CHAIR ROSENTHAL: Okay. So, let=s quickly vote on importance, 1, yes; 2, no.

How many people think this is important?

How many people think it is not important?

It=s unanimous.

Get the phone vote.

I could do that one because it was --

MS. FANTA: Importance, yes or no. Jeptha?

DR. CURTIS: Yes.

MS. FANTA: Doris?

DR. PETER: Yes.

MS. FANTA: And Ethan?

DR. HALM: Yes.

MS. FANTA: Okay. So, 15 yes, zero no.

CO-CHAIR ROSENTHAL: Okay. Good. Thank you. Enough time spent on that.

Now let=s do the scientific acceptability, reliability and validity.

Kurt?

DR. ELWARD: Yes. Yes, I think, overall, it was felt that reliability was moderate with a couple of highs. The measure, it does seem to identify claims that should be part of an episode of asthma, divided into year-long segments. I think, overall, we were satisfied that it had good reliability.

CO-CHAIR ROSENTHAL: Okay.

Questions either for Kurt or the TAP?

Jan, did you have anything you want to add to that?

DR. MAURER: No, I agree with Kurt=s statements.

CO-CHAIR ROSENTHAL: Any other discussion around reliability?

Jack?

DR. NEEDLEMAN: Yes, I have a question, just to clarify the measure. Are we talking the cost for people with asthma? Are we talking about the cost of asthma over a one-year period for the chronically-ill?

CO-CHAIR ROSENTHAL: Ingenix, can you clarify that?

MR. LYNN: It is the cost of asthma for a one-year period.

CO-CHAIR ROSENTHAL: With some index, diagnosis, or DRG submission that starts the episode, right?

MR. LYNN: Well, yes, you have to have diagnostic, you have to have a face-to-

face encounter between a clinician and a member with an asthma diagnosis.

CO-CHAIR ROSENTHAL: Okay. And then, it is one year?

DR. MAURER: And just to clarify that for other measures, there are some measures that allow a pharmacy claim as an initiating event or identification for asthma. This one does not.

MR. LYNN: No, that=s not true, not for us.

DR. MAURER: No, that=s what I=m saying.

MR. LYNN: Some people do that, but we don=t do it.

DR. MAURER: This does not, this particular measure.

MR. LYNN: Oh, I=m sorry. I=m sorry.

CO-CHAIR ROSENTHAL: Yes, she said for some types of episodes you could allow a pharmacy claim --

DR. MAURER: I=m trying to distinguish these.

CO-CHAIR ROSENTHAL: -- to start it, but this requires a face-to-face with a physician.

MR. LYNN: My apologies.

DR. MAURER: Yes, it does.

CO-CHAIR ROSENTHAL: Did somebody else over here have a --

DR. NEEDLEMAN: That=s inconsistent with what is in the TAP report for the description of this. So, can we get that clarified?

It says, ADescription@. AThis measure addresses the resource use of members identified as having asthma. Both encounter and pharmacy data are used to identify members for inclusion.@

CO-CHAIR ROSENTHAL: Oh, the fact that it says Apharmacy@ would start the episode, and that apparently is not correct. Okay. Right, it includes pharmacy claims

data, but a pharmacy episode does not start, does not initiate an episode.

Did somebody over here -- Steve? Any other comments on this, on the reliability?

(No response.)

Can you give us the TAP scores on this?

DR. RUDOLPH: I have one question.

CO-CHAIR ROSENTHAL: Yes, ma=am?

DR. RUDOLPH: Is the measure for all ages or is it specific to a certain age group?

CO-CHAIR ROSENTHAL: Ages? I assume it is up to 64, but what are the ages? Is it 18 to 64 or what are the ages?

MR. LYNN: I believe it is all ages with risk adjustment based on age.

CO-CHAIR ROSENTHAL: Okay. Barbara, are you all right with that? Okay.

And it is commercial. It is a commercial population.

So, can we get the TAP scores on reliability? And then, we will do our vote.

MS. WILBON: So 2a1, for well-defined, precise specifications, 2 high, 6 moderate, and 1 low. And reliability testing, 3 high, 5 moderate, and 1 low.

CO-CHAIR ROSENTHAL: And then, overall?

MS. WILBON: Overall was 8 moderate and 1 low.

CO-CHAIR ROSENTHAL: Okay. Heavy on the moderates.

If there is no further discussion, we are voting on overall reliability, and this is 1, high; 2, moderate; 3, low, and 4, insufficient.

(Whereupon, a vote was taken.)

MS. FANTA: Okay, and on the phone, overall reliability.

Jeptha?

DR. CURTIS: Moderate.

MS. FANTA: Doris?

DR. PETER: Moderate.

MS. FANTA: Ethan?

DR. HALM: Moderate.

MS. FANTA: Thanks.

CO-CHAIR ROSENTHAL: All right. Heavy doses of moderate.

MS. FANTA: So, we have 1 high, 14 moderate, and 1 low.

CO-CHAIR ROSENTHAL: Okay. So, now we will move to validity.

Kurt?

So, this is all the rest of the statistical stuff.

DR. ELWARD: Overall, the votes were moderate to high.

The determination of what is an actual asthma cost and what isn=t could have been more transparent. I think Ingenix tried to address this in the supplementary documents in a fairly good way. It is still difficult to sort out exactly what the programming is for this, but they responded that it involves

a number of markers, including diagnostic spirometry and exacerbation measures.

So, I think they tried very well to try to address the issue of validity to our satisfaction.

CO-CHAIR ROSENTHAL: Questions? Discussion?

Dolores?

MS. YANAGIHARA: I had a question. On top of page 15, it says that asthma with acute exacerbation is a condition status factor, and that the condition status factors are used to assign severity level.

It seems a little bit circular to me, if you are having an asthma exacerbation that is putting you into a higher severity level, which then you would expect a higher cost. Isn=t that what this is all about, managing asthma well? So, the exacerbation a symptom of not being managed well, and that is putting you into a higher severity level. It seems circular.

MR. LYNN: Yes, this is Tom Lynn from Ingenix.

Is that for the member, the developer?

MS. YANAGIHARA: I=m sorry, I didn=t hear what you said.

CO-CHAIR ROSENTHAL: He is asking is the question for them.

MS. YANAGIHARA: Sure.

MR. LYNN: Okay.

CO-CHAIR ROSENTHAL: Yes, the answer is yes.

MR. LYNN: Yes, you know, what we are trying to do is capture the cost of asthma and measure what are the markers that impact that cost. And the decision we made was, if it is diagnostic, then we should use it as a marker. If it is utilization directly, then we don=t.

I think what we are trying to do there, well, what we are trying to do there is it is possible that someone has an

exacerbation of asthma because they are poorly managed. But it is also possible that, you know, that is when the doctor gets the patient, is when they are poorly managed or they have a severe episode of asthma because of the initial diagnosis, and things like that.

So, we didn=t really feel like we could take it out of the marker because there are lots of situations where the doctor who ends up taking care of the patient wasn=t really, that his management or her management was not really the cause of the issue. So, we kept that marker in.

DR. ELWARD: I mean, one thing that I would just say, it is a huge challenge in general. If you look at the HEDIS and the NCQA measures, they are defined entirely on utilization. And despite a lot of efforts nationally at the NEPP to get even new diagnostic codes that say, if somebody has severe or persistent, you know, or

intermittent asthma, those don=t exist. There is sort of a CPT 2 code you can play with, but that is insufficient.

So, all across the board, everything related to severity is based on utilization, which, again, is circular.

DR. LEE: I think this is a good issue, and I think if you were asked, is it better to overadjust or underadjust, if you are going to err, I would vote for overadjusting.

(Laughter.)

DR. MAURER: I have one comment about this area, just reflecting what the conversation was at the TAP. There was some concern that pharmacy cost would not be adequately captured here, and that since they represent over 50 percent of the cost of managing asthma, that that might be a issue.

Maybe Ingenix would like to comment on that. Did they think they capture them better, or whatever?

MR. LYNN: Yes. No, I appreciate the opportunity to address that because it did come up at the TAP.

And basically, the point was that, hey, everything, they were talking about, has pharmacy information, but for asthma it is 50 percent of the cost; it is a bigger deal than for other things, was the point well-taken by the TAP.

What the grouper tries to do to deal with that, what the grouper does to deal with that is it says, you know, we can take a patient that has pharmacy benefit or does not have a pharmacy benefit. And then, we give a different -- then it is a different value than the expected value. We do this for all of the episodes. If you are a member that does not have pharmacy data, then you have a different expected cost than if you are a member that does have pharmacy data.

Now, having said that, there was some talk in the TAP that maybe for asthma you

should exclude the people that don=t have a pharmacy benefit, which is something that we would certainly consider, if that was the decision of the Steering Committee.

But it is corrected for.

DR. NEEDLEMAN: My question is a direct follow-on, but I would like to hear from the clinicians. I just heard 50 percent of the cost of asthma care is pharmacy. I am just wondering whether variations in pharmacy regimes, including potentially differences in the cost of the pharmacy regimes, are associated with the likelihood that you can keep the patient out of the ER, keep the patient out of the hospital.

Because it is not just enough to know whether it is excluded or included. If we are trying to understand how differences in resource use in one category affect resource use in the other, and we don=t have data in the category of interest, where variations exist and we think variations are important in

management, then I don=t see how we have got an adequate measure here.

So, that is a question to the clinicians. Are those premises about the role of pharmacy treatment and its impact on other costs that we want to look at correct?

DR. MAURER: You could certainly argue that. I mean it is fairly expensive for patients without coverage to buy inhaled steroids, which is the mainstay of people with persistent asthma. So, you could certainly argue that.

CO-CHAIR ROSENTHAL: Kurt?

DR. ELWARD: Well, maybe we need some more information from Ingenix. My impression was that they could separate out pharmacy, you know, look at pharmacy cost versus overall cost.

That is certainly important because, exactly, if I have people on -- given that all inhaled steroids are brand name and are charged as such, if I spend more money on

the pharmacy benefit for my asthma patients, I probably keep them out of the ER. So, yes, being able to look at those two different buckets of cost and say we know if you provide better asthma care, your pharmacy cost is going to go up, but your ER cost should go down.

CO-CHAIR ROSENTHAL: But this is the exact question that Jack posed in the last meeting, which was there are variable penetrants of availability of pharmacy cost. Isn=t that the point you have been making? And therefore, if you have got one group that has got pharmacy costs included and you try to compare it to a group where you don=t have the pharmacy cost, you are going to end up with incomparable figures.

DR. LEE: Yes, but here I am going one step further.

CO-CHAIR ROSENTHAL: Right.

DR. LEE: I=m saying the variation in pharmacy costs and our ability to drill

down on what we are doing in pharmacy in primary care has important information for helping us figure out how to improve our care. And if we don=t have that and we are not including it in our measure of resource use, we haven=t got enough information from our measure of resource use to help us figure out how to improve quality, how to improve care.

DR. MAURER: Yes, I think the TAP members who were discussing this would say that your inability to see where your costs are being expended in pharmacy or in other types of utilization might bias your interpretation of a measure like this, if you didn=t have accurate information.

DR. ELWARD: And this is Kurt.

I agree with Jack completely. I mean I think we tried to express this in the TAP, that there needs to be, if there is a differential access, then that needs to be made clear in any reporting of those measures.

CO-CHAIR ROSENTHAL: But, again, I

don=t think that that ever got specified, did it, in any of the other measures that we have looked at?

Jack, you would be the one who would remember this.

DR. NEEDLEMAN: Well, where I thought either the carved-out costs for pharmacy or mental health were going to be substantial, and where variations there might be influenced by the fact that there is a carve-out or not a carve-out, I choose to prescribe drugs because it is not in my risk pool, it is in somebody else=s risk pool.

I voted no because I didn=t think that the measure was complete enough, and I didn=t think the stratification on the basis of pharmacy costs, in the thing or not, were sufficient to enable the measure to be used to understand treatment decisions and the consequences of treatment decisions.

CO-CHAIR ROSENTHAL: Okay.

DR. NEEDLEMAN: I didn=t worry

about it on hip or knee. Most of that is hospital-based. We have got those costs included, and I assume the post-hospital drug regimes are fairly similar.

But this is one where I am very concerned that, if we don=t have the pharmacy data, we don=t have enough information --

CO-CHAIR ROSENTHAL: Got you.

Barbara?

DR. NEEDLEMAN: -- for it to be usable by the plans.

CO-CHAIR ROSENTHAL: I got you.

Barbara?

DR. RUDOLPH: At least in one part of the submission form it talks about the fact that they looked at what would cause the variation across providers, and that it was more likely to be things like referrals to esophageal specialists, hospitalizations, emergency department activity, those kinds of things that would actually create the larger variations among the provider groups. Now

maybe it is because they don=t have the pharmacy data in there.

But my feeling would be that you would see -- I mean, because the difference in cost between like a hospital stay and pharmaceuticals is, you know, pretty large. So, I would think that those things would pop the providers to a higher utilization than other things that are more routine but lower cost.

CO-CHAIR ROSENTHAL: But can you compare an entity, just on the face of it, that has pharmacy data with one that doesn=t?

DR. RUDOLPH: I couldn=t find that, but --

CO-CHAIR ROSENTHAL: And could they even sort out in their dataset the causes of variation, if some have pharmacy data and some do not?

DR. RUDOLPH: I think this was a study done, actually, by Weinberg, who looked at asthma.

CO-CHAIR ROSENTHAL: ***Yes, but what we are hearing here is that the pharmacy is driving half the cost, if that is accurate. I mean I am assuming that is accurate. That is what has been asserted.***

And, Jack, you have been consistent on this point. If we were to accept the premise that either mental health and/or pharmacy being variable as to whether it is reported at all, if it is a relevantly-sized or a material difference, or part of the treatment care, if we were to be consistent, we would say no to those where it is based on this methodology, and yet, those are big parts of the cost. And we might, then, be consistent in saying yes to others like hip and knee replacement, where the pharmacy costs are de minimis. That=s your point?

DR. NEEDLEMAN: ***Yes, that=s my point, and that I am trying to create measures, I want to make sure that we have measures that we can learn from, not just***

compare costs to.

So, if the pharmacy costs are carved out and they are invisible, and yet, the decisions that are being made in pharmacy therapy, you know, the drug therapies for patients, are making a big difference in their risk of being in the ER, being admitted to the hospital, and there are systematic differences in prescription patterns because in some cases my plan owns those costs, in other cases the carve-out folks own those costs, so we are making different decisions, all that is invisible. And therefore, we can=t learn from that experience.

CO-CHAIR ROSENTHAL: Okay. Barbara, and then Paul.

DR. RUDOLPH: In the specifications, to those who create the data for this measure, it says, AA member=s pharmacy benefit status should be noted and reflects whether or not the member has pharmacy data generally available for use in

measurement. It is recommended for this measure that members without continuous pharmacy benefit be excluded from the asthma resource use measure. Examples of populations where pharmacy data may not be available include the individual who does not have pharmacy coverage for the defined enrollment period of pharmacy services managed by the PDM and the PDM....@

So, they are pretty specific about who to include or not include in this.

DR. REDFEARN: Okay. So, they account for that and say only compare apples to apples. All right. Okay.

CO-CHAIR STEINWALD: More than that, the apples have to have pharmacy --

(Laughter.)

DR. NEEDLEMAN: Well, those are the apples to apples. That=s what I meant.

CO-CHAIR ROSENTHAL: No, I=m just trying to clarify. Okay.

MR. LYNN: But, Cheri, correct me,

at this time in Ingenix, correct me, is that an edit that would just take into account the TAP comments?

CO-CHAIR ROSENTHAL: Would you repeat that?

MR. LYNN: So that, for asthma, we are excluding members that don=t have a pharmacy benefit?

DR. MAURER: Yes, we made that modification in the document that we sent, the Word documents.

MR. LYNN: Okay. I=m sorry. I had forgotten that we had done that. I apologize.

CO-CHAIR ROSENTHAL: Okay. So, now Ingenix has clarified that for themselves.

Paul?

MS. ZIELINSKI: Let me double-check that, but I am pretty sure that we did make that change.

CO-CHAIR ROSENTHAL: Barbara is reading it right out of something.

DR. RUDOLPH: It=s on page 12 of the submission form.

MS. ZIELINSKI: Oh, we made a modification, and we sent that to Ashlie on August 11th.

CO-CHAIR ROSENTHAL: She=s talking to them. They are clarifying it internally, I think. They are talking among themselves.

Paul?

DR. BARNETT: Yes, so I just --

MS. ZIELINSKI: I=m letting you know we had a modification to the submission that was sent to the NQF on August 11th.

MR. LYNN: Yes, Cheri, and they have that.

MS. ZIELINSKI: Oh, okay.

MR. LYNN: Yes.

CO-CHAIR ROSENTHAL: Okay.

DR. BARNETT: So, I just wanted to clarify, thinking again about that question about the utilization driving the risk factor, that if someone has an emergency visit or

hospitalization for asthma, so is that in the current period, the one that you are adjusting, or is it in some prior period that you are using to make that adjustment?

MR. LYNN: No, let me make that clear. We are not -- in no place is an emergency room visit used as a severity marker. That is the utilization. We don=t use that, whether it happened before the episode or during the episode, we don=t use that as a severity marker.

DR. BARNETT: Well, exacerbation it was.

MR. LYNN: Yes, that is a diagnostic, and we do use that, and we use it when it occurs during the episode.

DR. BARNETT: But isn=t an exacerbation likely to result in emergency room utilization? I mean that is where the code is going to get assigned, right?

MR. LYNN: Well, we are not using utilization directly.

DR. BARNETT: Yes, but --

MR. LYNN: I mean, to the extent the diagnosis -- the utilization, I mean, that is what we are trying to -- that is why it is a severity marker.

DR. BARNETT: Okay. So, let me rephrase the question then. Is it exacerbation in the current period that would affect the risk factor or is it an exacerbation that occurred in a prior period?

MR. LYNN: It=s the current period.

DR. BARNETT: Yes, so it seems, since that is so tightly linked with utilization, it seems to violate one of the principles of risk adjustment. So, suppose that a clinician does a really terrible job and all of the patients have exacerbations. Then, all of their patients have high cost. But because we adjust for this in the risk factor, this looks like an efficient provider, the one who everybody has an exacerbation.

MR. LYNN: I guess our decision was to err on the side of risk-adjusting for --

DR. BARNETT: Well, so I think the proper way to deal with this is, did the patient have exacerbation in a prior period? That would mean that they were at high risk in this period, and that would be an appropriate case mix measure that doesn=t reflect the management in the current period.

But to use the outcome as a case mix variable is not good.

CO-CHAIR ROSENTHAL: Okay. And I have one question. From the TAP discussion, on the piece of paper we have, it does say here, ATo examine how refined the risk adjustment is, R-squareds for different severity levels and how they predict resource utilization should be provided.@

For the Ingenix people, did this request actually make it to you all or not?

MS. ZIELINSKI: I=m sorry, what

was the request again?

This is Cheri.

MR. LYNN: It is the R-squared for asthma. Did we get a request for R-squared for asthma?

DR. ELWARD: No, I=m sorry, a little bit farther down. Yes, page 4.

CO-CHAIR ROSENTHAL: So, do we have the answer on this one?

DR. ELWARD: Yes.

CO-CHAIR ROSENTHAL: Kurt, can you help us?

DR. ELWARD: They actually talk about, I mean, they have a few different R-squareds for hospital admissions, stays per episode, ER visits, specialty visits, pharmacy scripts. And they range from 0.5 to 0.9.

CO-CHAIR ROSENTHAL: Okay. All right.

DR. ELWARD: Yes.

CO-CHAIR ROSENTHAL: Thank you.

Any other questions, discussions,

on overall validity?

DR. ELWARD: Yes, this is Kurt.

I would agree with the last comment. I wasn=t aware that they were adjusting within the period. So, I think that is a very important comment.

Also, just for clarification, I think it is on page 12, as Barbara mentioned, they talk about the pharmacy benefit status and say, if members without continuous pharmacy benefit -- they recommend that members without continuous pharmacy benefit be excluded. So, I guess that is the closest they get to it. But I would say, clearly, that needs to be, pharmacy claims, as Jack said, have to be included in the model.

CO-CHAIR ROSENTHAL: Right. Or excluded, so you are comparing apples to apples.

DR. ELWARD: Or make it very, very clear, yes.

CO-CHAIR ROSENTHAL: Well, I=m

sorry. If somebody doesn=t have pharmacy benefit, they would be excluded from the analysis. In the analysis would be people with pharmacy benefit. So, you are comparing apples to applies.

Okay. Hearing no further discussion, can we get a tabulation of the TAP scores? And then, we will call the question on overall validity.

MS. WILBON: All right. So, for the subcriteria for validity, 2b1, the specifications are consistent with the resource use or cost problem. We have 2 high, 5 moderate, 1 low, and 1 insufficient. For validity testing, we had 1 high, 4 moderate, and 2 low. For exclusions, 1 high, 7 moderate, and 1 low. For risk adjustment, 1 high, 4 moderate, 2 low, and 2 insufficient. And for 2b5, identification of statistically-significant, meaningful differences, 8 moderate.

MR. AMIN: Just for consistency

purposes, 2b1, again, does not reflect the changes in the costing approach.

CO-CHAIR ROSENTHAL: Okay, and there is a comment in the TAP saying, though, their concerns about it not being standardized pricing, yes.

And then, overall?

MS. WILBON: And then, right, overall validity was 6 moderate, 1 low, and 2 insufficient.

CO-CHAIR ROSENTHAL: Okay. So, I think we have the TAP report. We have had a thorough discussion on this. So, our vote will be on overall validity. One, high; 2, moderate; 3, low, and 4, insufficient.

Sarah, turn this on.

(Whereupon, a vote was taken.)

MS. FANTA: Okay, and on the phone, for overall validity.

Jeptha?

DR. CURTIS: Moderate.

MS. FANTA: I=m sorry?

DR. CURTIS: Moderate.

MS. FANTA: Okay. Thanks.

Doris?

DR. PETER: Moderate.

MS. FANTA: Okay. Ethan?

DR. HALM: Moderate.

MS. FANTA: Okay. So, we have --

CO-CHAIR ROSENTHAL: Reluctantly?

(Laughter.)

Maybe not so reluctantly this time.

MS. FANTA: So, we have 8 moderate and 8 low.

CO-CHAIR ROSENTHAL: All right. And now we need to overall scientific acceptability, if there is no further discussion. So this now is 1, yes; 2, no.

(Whereupon, a vote was taken.)

MS. FANTA: And on the phone, overall scientific acceptability.

Jeptha?

DR. CURTIS: Yes.

MS. FANTA: Doris?

DR. PETER: Yes.

MS. FANTA: Ethan?

DR. CURTIS: No.

MS. FANTA: Okay. So, we have 8 yes and 8 no.

(Laughter.)

CO-CHAIR ROSENTHAL: We quit. We quit. I am going to speak for Bruce.

(Laughter.)

Helen, this one is, obviously, a complete split decision. Shall we do usability and an overall? Let=s just finish it up.

DR. BURSTIN: I believe there is a competing measure you are going to have shortly. So, it would be nice to have this. Well, they are different levels. You are going to talk about asthma shortly again. It would be nice to finish it up.

CO-CHAIR ROSENTHAL: Okay. So, let=s quickly discuss usability. And again,

unless there is something terribly different about this than the, say, pneumonia measure, et cetera, I am assuming we won=t need a ton of conversation or questioning about this.

DR. ELWARD: No, I would say the comments are about the same. It felt like this was probably more usable than the pneumonia measure.

CO-CHAIR ROSENTHAL: Okay. Steve?

MR. PHILLIPS: Yes, I guess my only question was on the length of the episode, and in looking at it, from what I could find, it is recommended that there be a one-year window.

I guess it would seem to me preferable to make that part of the specification because, if we are endorsing the measure but users are able to use an alternative episode, I would have some concern about that.

CO-CHAIR ROSENTHAL: Ingenix, can we get some clarification on that? I thought

that, in fact, it specified specifically that it was one year.

MR. LYNN: Yes, the intention was, it is specified for one year.

CO-CHAIR ROSENTHAL: Steve, is there some language there that you are referring to that would call that into question?

MR. PHILLIPS: Yes. One second. Okay, yes, I=m looking at page 21. In terms of episode completeness, asthma is a lifelong condition. I guess the last sentence there in parentheses, AFor the convenience of analytics and measurement, it is customary to segment chronic episodes, including asthma, into year-long episode units.@ And I may have missed it, but I was just looking for a more definitive statement that the measure should be --

MR. LYNN: Yes, I think that sentence was meant to defend the idea of dividing it into year-long episodes, but the

specification is year-long episodes.

CO-CHAIR ROSENTHAL: All right. I think we are hearing clarification that the answer is, yes, that is the specifics on it.

DR. ELWARD: And as you think through it, I mean I think their logic is -- I am not sure that asthma should be thought of as episodes because it is a chronic condition, and what you want to do is actually decrease episodes of acute care. But I think their rationale makes sense.

CO-CHAIR ROSENTHAL: Yes. Well, it is called an episode because it is called an episode grouper. So, you have to call it an episode. But, anyway, semantics.

Okay. Any further discussion on this point?

(No response.)

I know everybody wants a break here desperately.

So, this is overall usability. It is 1, high; 2, moderate; 3, low; 4,

insufficient.

And can we just get the TAP score on this real quickly? Ashlie?

MS. WILBON: Sorry. 3a is the performance results are publicly reported. Two high, 4 moderate, 2 low, and 1 insufficient. The measure results are meaningful and useful for public reporting and performance improvement. That is 3b. Six moderate, 2 low, and 1 insufficient. And 3c, the data results can be deconstructed for transparency and understanding, 3 high, 5 moderate, and 1 low.

CO-CHAIR ROSENTHAL: Okay. So, we=re 1, high; 2, moderate; 3, low; 4, insufficient.

Sarah?

(Whereupon, a vote was taken.)

MS. FANTA: And on the phone, for overall usability.

Jeptha?

DR. CURTIS: Insufficient.

MS. FANTA: Doris?

DR. PETER: Moderate.

MS. FANTA: Okay. And Ethan?

DR. HALM: Low.

MS. FANTA: Okay. So, we have 9 moderate, 6 low, and 1 insufficient.

CO-CHAIR ROSENTHAL: All right. And we will consider the feasibility score to be unchanged.

And the last item that we need to do as a group on this measure is recommendation for endorsement overall. So, 1, yes; 2, no; 3, abstain.

Any further discussion before we do overall recommendation for or against endorsement?

(No response.)

All right, hearing none, Sarah?

(Whereupon, a vote was taken.)

MS. FANTA: And on the phone, overall recommendation.

Jeptha?

DR. CURTIS: Yes.

MS. FANTA: Okay. Doris?

DR. PETER: Yes.

MS. FANTA: Okay. Ethan?

DR. HALM: No.

MS. FANTA: Okay. So, we have 7 yeses and 9 noes.

CO-CHAIR ROSENTHAL: All right. This concludes discussion on this measure.

I think we will take a quick break and then resume and finish up.

Yes, Paul?

Oh, you=re just shielding from the sun?

(Laughter.)

Fifteen minutes.

MS. WILBON: For those on the phone, it is about three o=clock. We will reconvene at 3:15.

Thank you.

(Whereupon, the foregoing matter went off the record at 2:57 p.m. and resumed

at 3:18 p.m.)

CO-CHAIR ROSENTHAL: Okay, we=re now going to do 1608.

MS. ZIELINSKI: Excuse me. I apologize for interrupting. This is Cheri Zielinski from Ingenix.

CO-CHAIR ROSENTHAL: Yes, ma=am?

MS. ZIELINSKI: I know that before the break you had mentioned that there was going to be some discussion on consistency with the voting. Are we going to be privy to those discussions at all? Or I am just wondering what the outcome of those discussions was.

MS. WILBON: We have not had that discussion yet, Cheri.

This is Ashlie. Hi.

We are going to finish this last Ingenix measure, and then we are going to discuss, we will probably discuss when and how we should have that discussion. So, we haven=t had it yet, though. And I do believe

that would be open to the public as well. That would be open.

MS. ZIELINSKI: Thank you.

CO-CHAIR ROSENTHAL: But I think we want to get through the remaining measures --

MS. WILBON: Yes.

CO-CHAIR ROSENTHAL: -- this afternoon.

MS. WILBON: Yes, we need to at least get through the last Ingenix measure. And then, we will decide when to have that discussion.

CO-CHAIR ROSENTHAL: Yes, yes.

MS. WILBON: Okay?

CO-CHAIR ROSENTHAL: Okay. So, 1608 is open now, and this is the COPD cost-of-care measure for Ingenix.

So, Kurt?

DR. ELWARD: It=s a problem.

CO-CHAIR ROSENTHAL: Okay. So, can we quickly vote on importance?

All who believe that COPD is an important measure to be dealing with, raise your hand.

Any opposed?

Anybody on the phone believe this is not important?

(No response.)

Okay. So, let=s, then, move right to the scientific acceptability, and I think in doing so, what is either similar about COPD or different from pneumonia and asthma can be featured in the discussion.

So, Kurt, do you want to start us off on -- now we will talk scientific acceptability?

DR. ELWARD: Yes, overall, there were medium to high levels of the reliability.

We did raise questions around the timeframe. Initially, that was 180 days. Ingenix, subsequently, responded that that will be a year also, consistent with the asthma measure.

And we felt that the results were repeatable.

The overall reliability was felt to be high to moderate.

CO-CHAIR ROSENTHAL: Okay. Sorry, I think a couple of us are hunting through our stuff to be sure we have the right piece of paper.

Open for discussion then.

(No response.)

Questions or comments?

(No response.)

Any differences that are substantive from -- I would say this is more like the asthma discussion in that this is a chronic disease, and the measurement period is one year in length.

MR. BOWHAN: How prevalent is it among under-65s?

DR. ELWARD: It is still significantly prevalent, say, over 45.

CO-CHAIR ROSENTHAL: Yes, I think

it is pretty prevalent.

***DR. RUDOLPH:** And misdiagnosed in younger people. Generally, they are given an asthma diagnosis instead of COPD.*

***CO-CHAIR ROSENTHAL:** Is that a relevant factor then to the question about reliability if, in fact, it is misdiagnosed frequently?*

***DR. RUDOLPH:** In young people.*

***CO-CHAIR ROSENTHAL:** Only in young people? You mean like ages 18 to 64, for which the measure is -- okay. I am exaggerating that for effect, but, I mean -- yes, sir?*

***DR. BARNETT:** I just want to ask, is the same issue with exacerbations part of the case mix measure, as was true in the asthma measure?*

***CO-CHAIR ROSENTHAL:** Tom, is the exacerbation issue the same as it was in asthma?*

***MR. LYNN:** Yes, we are looking at*

markers that are for during the episode.

DR. BARNETT: I=m sorry, could you repeat that? It wasn=t quite clear. You=re looking at markers of?

MR. LYNN: That occur during the episode. This works the same as asthma.

CO-CHAIR ROSENTHAL: Okay. So, his answer is it sounds like it is the same.

DR. ELWARD: Tom, maybe you could explain a little bit more because I wasn=t aware of that. And can you explain what the rationale has been for using it that way?

CO-CHAIR ROSENTHAL: Tom, did you hear the question?

MR. LYNN: Again, the rationale is that we don=t want to -- we are more concerned then about identifying the physician who picks up a case with COPD exacerbation as a new provider for that member, and not adjusting in that case, than we are about making sure that we don=t adjust in the case where the cause is mismanagement. A lot of times the cause is a

new diagnosis or it is an episode where a member ends up going to another doctor. So, that is why we made that decision.

DR. ELWARD: How do you get around the adjustment, the issues, though, that have been mentioned in terms of sort of one feeding into another? On the one hand, it could be that, I mean, there is credit in assigning resources to poorly-managed patients because those exacerbations, if they are not managed well, should accrue to that provider or that institution. On the other hand, they could be reflective of more severe disease.

Is there something within your program that addresses that or tries to factor that in?

MR. LYNN: All we can do is look at the diagnostic information, and we can make decisions about whether to do things during the episode or prior to the episode, but we are looking at things that occur during the episode.

And the reason, it is not a statistical thing; it is a clinical thing. You know, it is probably more frequent that these are -- and I am not a pulmonologist; my training is in family medicine -- but, you know, it is probably more frequent that these are new cases or new to that doctor that have these sorts of exacerbations and not cases where they are poorly managed.

CO-CHAIR ROSENTHAL: Other questions or comments?

DR. PETER: Just a question -- Doris -- about the pharmacy, whether it is handled the same way as the asthma measure.

MS. ZIELINSKI: Yes, I can answer that. It is, I believe, but it is important to recognize in COPD that pharmacy is a much lower percentage of the cost of care than it is in asthma.

DR. PETER: It is more, I guess, a third or something, right?

MS. ZIELINSKI: Twenty percent I

think.

DR. PETER: It=s 20 percent?

MR. LYNN: We did not make the exception for COPD that we made in asthma. We rewrote asthma to exclude people that didn=t have a pharmacy benefit. We did not do that with COPD.

CO-CHAIR ROSENTHAL: All right. And the reason there is that the pharmacy costs are not as significant a component of the cost of care for COPD as they were for asthma?

MR. LYNN: That=s correct.

CO-CHAIR ROSENTHAL: Okay. Thank you for that clarification.

Any other questions or comments?

(No response.)

All right. So, I think we are ready to talk about 2a. So, if we could see the TAP scores? And then, I might suggest, is it possible, Ashlie, that we can see our vote on asthma? Or remind us of our vote?

MS. WILBON: Yes.

CO-CHAIR ROSENTHAL: In light of this question about -- and it doesn=t mean we have to be consistent. We clearly could say, no, no, no, COPD is really different and I=m changing my vote. But I haven=t heard an awful lot that is different, and it might be nice to at least see what we did 20 minutes ago at the point at which we vote. So, is it possible you guys -- you don=t have to show it on the screen, but you can tell us. Yes?

DR. BARNETT: Fourteen medium, 1 high, 1 low.

CO-CHAIR ROSENTHAL: Okay. That was the reliability vote on that. Okay, perfect.

And then, give us the TAP quickly on reliability.

MS. WILBON: Okay. So, for reliability, 2a1, about whether or not the specifications are precisely defined, 4 high, 3 moderate. Reliability testing, 5 high, 2

moderate.

CO-CHAIR ROSENTHAL: Okay. Perfect. And then, overall?

MS. WILBON: Overall reliability, 4 high, 3 moderate.

CO-CHAIR ROSENTHAL: Okay. Everybody prepared to press their clicker? So, for us, it is 1, high; 2, moderate; 3, low; 4, insufficient.

Point at Sarah starting now.

(Whereupon, a vote was taken.)

MS. FANTA: Okay, and for those of you on the phone, overall reliability, high, moderate, low, or insufficient.

Jeptha?

(No response.)

Doris?

DR. PETER: Moderate.

MS. FANTA: Ethan?

DR. HALM: Moderate.

MS. FANTA: Okay. And Jeptha, are you there?

(No response.)

So, we have 3 high, 10 moderate, and 2 low.

CO-CHAIR ROSENTHAL: Okay. Let=s now discuss validity.

Kurt?

DR. ELWARD: Overall, the validity was felt to be moderate to high in terms of consistency with intent.

They scored more moderate in terms of our concerns about the method for customization and the inability to compare actual versus standard prices. Now I think it was done this morning; they have chosen to change that to actual prices, so that I think we would probably rank that a little bit higher, certainly no worse.

There was a challenge in sort of the tiebreaking logic and how, if you weren=t sure -- and maybe, Janet, you can help me out with this -- about how they actually, given the number of comorbidities that COPD patients

you have, how you break that tie in terms of, if you are not sure whether or not it relates primarily to COPD or the patient=s heart failure.

DR. MAURER: Yes, exactly. I mean many COPD patients have accompanying heart disease because it is the same underlying cause. And heart failure versus an exacerbation becomes a real difficult differentiating factor. So, where do you put it?

The other thing that was brought up around COPD and severity, the severity score is done in a similar way to the asthma score. But people who take care of COPD are more used to thinking of mild, moderate, severe COPD in terms of the amount of lung dysfunction rather than the comorbidities as much. So, there was some discussion around that.

But, in the end, you know, it was more focused around, of the comorbidities,

which is more important, and which one do you end up in; which category do you end up in, cardiovascular or COPD or where?

CO-CHAIR ROSENTHAL: Other questions, comments?

Yes, Jack?

DR. NEEDLEMAN: At the risk of sounding like a broken record -- (laughter) -- I am looking at the supplementary materials provided, and looking particularly at Table 1 and Table 2. And Table 2 is the average cost across all the severity categories for the different categories of cost. And 33 percent of the costs of the COPD patients are in pharmacy in every severity category. That is Table 2. And in Table 2, it is the second largest cost after hospitalization, which is 34 percent of the cost.

And if you look at Table 1, in every severity category except the highest, pharmacy costs are the largest single category, far exceeding any other cost,

including hospitalization, on average.

In the lowest severity category, there are four and a half scripts per patient on average in this category. I do not understand how we can understand resource use without understanding pharmacy use.

CO-CHAIR ROSENTHAL: So, does that jibe with what was stated, that pharmacy costs are not a significant component of COPD?

DR. NEEDLEMAN: Well, you know, they weren=t as significant as asthma.

CO-CHAIR ROSENTHAL: Okay.

DR. NEEDLEMAN: I=m telling you how significant they are without comparing to asthma. I find these incredibly significant.

And if we are trying to understand variations in resource use, we have got to understand variations in pharmacy use. And you can=t do that if you haven=t got the pharmacy data.

CO-CHAIR ROSENTHAL: Any other comments on that?

I would also say I have jumped ahead, but, again, a little bit of notion of being consistent in that the COPD measure for NCQA, the TAP discussion talks about one of the challenges that COPD has multiple comorbidities, particularly when compared to asthma, and it will be difficult, therefore, to know if you are measuring exactly COPD. So, that observation was made for the NCQA measure. I believe it would also have to apply similarly to this one because the issues are exactly the same, unless I am missing something.

DR. MAURER: Well, I think there is a difference, actually. NCQA doesn=t even begin to say that they are trying to measure just the cost related to COPD. They are saying that they are measuring all the costs that a patient with COPD had in that measurement year.

This is more attributing the cost to a specific disease. I think that is where

the difference.

CO-CHAIR ROSENTHAL: Okay. So, the difference is that the other costs are excluded from this one?

DR. MAURER: Yes.

CO-CHAIR ROSENTHAL: Only COPD-related costs --

DR. MAURER: They=re supposed to be.

CO-CHAIR ROSENTHAL: -- are included?

DR. MAURER: Yes.

CO-CHAIR ROSENTHAL: All right. Well, that is an important distinction.

DR. MAURER: But the question would be, how do you actually figure out what to exclude and what not to, you know?

CO-CHAIR ROSENTHAL: Right. That would be a question.

Okay. Are there other observations, questions, or comments about overall validity?

(No response.)

All right. Can we see what the TAP votes were and what we said about asthma?

MS. WILBON: So, I was trying to bring up the asthma votes.

CO-CHAIR ROSENTHAL: Okay. Oh, zero high -- well, let=s do the TAP --

MS. WILBON: Okay.

CO-CHAIR ROSENTHAL: -- and then we will do our previous vote on asthma.

MS. WILBON: So, the TAP votes for validity, 2b1, whether the specifications are consistent with the resource use or cost problem, is 2 high, 5 moderate. Validity testing, 7 moderate. Exclusions, 1 high, 6 moderate. Risk adjustment, 4 moderate, 3 low. And identification of statistically-significant and meaningful differences, 7 moderate.

CO-CHAIR ROSENTHAL: All right. And then, David, what was our vote on asthma?

Zero high, 8 medium, 8 low, zero

indeterminate, okay, or insufficient.

Okay. Yes, sir, Paul?

DR. BARNETT: Just a question. So, if we think this might conflict with another NQF-endorsed measure, where does that fit in the taxonomy of things we consider here?

MS. WILBON: Well, as we review each measure, before we even get to kind of whether or not it conflicts or is the same, we review each measure individually on their own merits. At the end, if you guys decide you want to recommend it, then we kind of look at what has been recommended as a pile and decide which ones are similar and which ones --

DR. BARNETT: No, I mean one that has already been endorsed in the past for quality measures.

MS. WILBON: Well, there haven=t been any -- oh, quality measures?

CO-CHAIR ROSENTHAL: Well, let=s find out what he means by Aconflicts with@

first --

MS. WILBON: Yes.

CO-CHAIR ROSENTHAL: -- before we try to answer it.

DR. BARNETT: So, the issue of what I discussed before about the current exacerbation triggering, being considered in the risk factor. It seems like it offers an incentive to not be concerned about ambulatory-sensitive hospitalizations. So, hospitalization for COPD is one of the ambulatory-sensitive conditions. The good primary care physicians keeps their patients out of the hospital.

So, here we are risk-adjusting for that. It seems like I guess it is one of those unintended consequences.

CO-CHAIR ROSENTHAL: Well, I think that is where it would have to be factored in our scoring of this. If we believe there is -- and I can=t remember where that -- is that in the usability part, unintended

consequences?

MS. WILBON: It=s in usability.

CO-CHAIR ROSENTHAL: Yes, it=s in usability.

MS. WILBON: Yes.

CO-CHAIR ROSENTHAL: It could be, it would be or could be a relevant factor in that vote, I think is the answer.

MS. WILBON: It=s actually feasibility. Sorry.

CO-CHAIR ROSENTHAL: Okay. Well, it=s in there somewhere. It=s in there somewhere. That=s where you would consider it.

Okay. So, we have our history on this. We have our TAP vote.

I=m sorry. Use your microphone.

DR. RUDOLPH: So, wouldn=t it be in validity because we are discussing, he is discussing a risk-adjustment factor?

CO-CHAIR ROSENTHAL: Well, I guess if you believed it was an inappropriate risk-

adjusting factor, you could vote here. If you thought it was an appropriate risk-adjusting factor for the cost, and yet, created an unintended consequence on the quality side, it would be voted in feasibility.

So, you know, I think we are splitting hairs, but I created the hair-splitting thing. So, I am forced to apologize for that, yet again.

Okay. Is there anything further?

(No response.)

I would say we should vote. One, high; 2, moderate; 3, low, and 4, insufficient, and we are voting 2b, overall validity.

(Whereupon, a vote was taken.)

MS. FANTA: Okay, and on the phone, overall validity, high, moderate, low, or insufficient.

Jeptha?

(No response.)

Doris?

DR. PETER: Moderate.

MS. FANTA: Thanks.

Ethan?

DR. CURTIS: Moderate.

MS. FANTA: Okay. So, we have 1 high, 5 moderate, and 9 low.

CO-CHAIR ROSENTHAL: All right. So, now we vote overall scientific acceptability, and this is yes or no.

Yes, you might as well, yes, give us what we did on asthma just again, so we know it.

MS. WILBON: Asthma was actually split 8 yes and 8 no.

CO-CHAIR ROSENTHAL: Okay. So, asthma was 8 yes, 8 no, for what that is worth. You are not bound by that in any way, shape, or form. This should be voted on entirely on its own merits.

But 1 is yes and 2 is no.

(Whereupon, a vote was taken.)

MS. FANTA: And on the phone,

overall scientific acceptable, yes or no.

Doris?

DR. PETER: Yes.

MS. FANTA: Okay. Ethan?

DR. CURTIS: Yes.

MS. FANTA: Okay. Great. Thanks.

So, we have 3 yes and 12 no.

CO-CHAIR ROSENTHAL: No, 5 yes.

MS. FANTA: Oh, 5. Sorry.

CO-CHAIR ROSENTHAL: Yes, you=ve got to add that.

MS. FANTA: Yes, 5 yes and 10 no.

CO-CHAIR ROSENTHAL: All right. So, we=re done, okay, with that measure and with the Ingenix measures.

Now, as a point of order, are the NCQA people prepared to start?

(Laughter.)

Touchdown. Touchdown. Sorry, I didn=t see. I didn=t see.

So, the suggestion is being made that we now have a brief, or as long as it

takes, discussion about whether we have been internally consistent with the various Ingenix measures, given that we rejected several of them and we approved the asthma measure, right? It was asthma that we approved?

MS. WILBON: We did a quick graph of how you guys have voted on all the Ingenix measures so far.

CO-CHAIR ROSENTHAL: Oh, that=s right. Okay. So, yes.

MS. WILBON: We didn=t do the COPD one that we just voted on.

So, the green, obviously, is the yes votes, and the red is the no votes. So, the square around on the right that you see, those are the four measures that you guys revoted on this morning in the context of that costing discussion.

CO-CHAIR ROSENTHAL: Okay. All right, but the three that, as of this moment, we have approved are the 12-to-4, the 9-to-7, and the 9-to-7. And which ones are they?

Well, just tell us what they are.

MR. AMIN: Can I just go systematically from the left to the right?

CO-CHAIR ROSENTHAL: Sure.

MR. AMIN: Okay. So, from the left is the ETG asthma measure, 9 to 7. COPD is skipped over. ETG pneumonia is --

CO-CHAIR ROSENTHAL: Well, move the marker there as you are doing it, if you would do that.

MR. AMIN: Okay.

CO-CHAIR ROSENTHAL: There we go.

MR. AMIN: There we go. This one right here is pneumonia, 12 to 4. Hip fracture is 9 to 7, hip and knee, 9 to 7. Non-condition-specific, 5 to 9. Diabetes, 7 to 7. CHF, 6 to 8; yes, 6, 8 no. And CAD, 5 yes, 9 no.

CO-CHAIR ROSENTHAL: I would like to suggest that that is actually pretty internally consistent. The three that were approved pretty, either overwhelming in the

one case or by a close vote in the other two, are kind of condition-specific where the marker of a starting point and a stopping point, again, somewhat intuitively hangs together.

Well, the diabetes, but it is 7 to 7. Yes, that is diabetes. The 7-to-7 one was diabetes.

MS. ZIELINSKI: So, this is Cheri.

So, CHF is not considered a condition-specific or CAD?

CO-CHAIR ROSENTHAL: No, no. Well, I am not going to argue it or debate it. I am just giving my own perception of it, that the three seem to me to make sense. Either a hip fracture or pneumonia is an acute event that has a starting point that most people can go, AOh, I get that.@ Even the attribution, whom is probably responsible in the case of a fracture or a knee replacement, it is the orthopedist that does the case. So, there is no debate about that. And those were

approved.

The others are chronic diseases. And the only one -- and it didn=t pass, but it is 7 to 7 -- is diabetes. And that seems to me the only one that is somewhat consistent, but that is just my read of the thing.

I would open it up for discussion.

MS. ZIELINSKI: This is Cheri.

So, didn=t asthma pass 9 to 7?

MS. WILBON: So, excuse me, Cheri. I need just a point of order.

So, I think our whole reason for wanting to do this in the context of the discussion of this morning was more around them changing their measures from using both standardized prices to actual prices or actual prices paid. So, we just want to make sure that, in the context in which you made those votes, if you voted down those three of the four measures because of that, if that is something that carries over into other measures, then that should be reflected.

If not, that=s fine, but we didn=t have a detailed discussion of those four measures because you have already had that. So, that revote seemed to reflect your feelings about or your sentiment about having actual prices only.

If that is not the case, that is fine, but we just want to clarify that, to make sure that the reason for voting those four measures down is consistent with --

CO-CHAIR ROSENTHAL: I think what Jack said this morning, though, is correct. It only took one or two vote changes to shift those votes from being positive to being negative. And it would argue that the fairly strongly positive votes on these three are in knowing that it is priced, that people took that into account in making these positive votes.

But, again, I am guessing at people=s motivation a bit. But I am assuming people took that into consideration as we made

the afternoon votes.

Paul?

DR. BARNETT: Cheri was confused about the vote on the asthma. It is 7 yes, 9 no. So, the asthma did not pass.

CO-CHAIR ROSENTHAL: Are there any other comments on the two aspects, what I guess are two aspects of some notion of consistency here?

(No response.)

I think we did a damned good job, frankly, I mean given the complexity of this.

But, Helen?

DR. BURSTIN: Yes, actually, I would just point out, it is very interesting, we went through a similar exercise last year and looked at the avoidable complications measures that submitted by Prometheus. It was the acute conditions that actually did well as well, and the chronic conditions that got all fuzzy that did not make it through, interestingly enough, except for an overall

one of all avoidable complications went through as well, but not for the chronic conditions.

So, you actually might be pointing out -- you know, there is a little bit perhaps more specificity and comfort around the attribution rules perhaps around those conditions, the acute ones.

CO-CHAIR ROSENTHAL: Barbara?

DR. RUDOLPH: Well, that may be true, but, I mean, I am really concerned because the money in this country is being spent on chronic care, and we are not doing our job here, or whatever, if we are not having any measures go through, measure resource use, about chronic conditions. Are we part of the problem?

DR. HALM: Well, this is part of the challenge with the episodic approach.

CO-CHAIR ROSENTHAL: Paul, do you want to weigh-in on this?

DR. BARNETT: Well, just to

observe that we have two left. So, hold your fire there. We=ve got two more measures.

And I think there is also a different approach in terms of whether we try to attribute cost to an episode or look at some larger group of costs and then control for case mix in that method.

CO-CHAIR ROSENTHAL: Right.

DR. BARNETT: So, we will see. Maybe we will have some more things to endorse.

CO-CHAIR ROSENTHAL: I would also say, from my point of view, I would have changed all my votes had the attribution not been at the individual physician level. If you attributed these to groups, any size group, relative size group, I probably would have changed my vote on several of them.

Your point about the cost being in chronic disease is well-made, but our job is to try to adjudicate these against science and whatnot, and maybe more work needs to be done

in those. And hopefully, more work will be done.

But I also agree with Paul; let=s hold our fire. We=ve got a couple more.

But are we satisfied that we have met any sort of hurdle or threshold for some level of consistency, without revoting?

How many people want to revote?

(Laughter.)

I could force the question that way quick quickly, couldn=t I?

Kurt?

DR. ELWARD: But, Tom, I think a couple of things come to mind, and this might be helpful for Ingenix. The thing I am hearing is that, if there are certain enhancements made, such as being able to do a little bit different approach to risk adjustment, making sure pharmacy benefits are included, I mean those two things would really have opened up -- oh, and aggregating at the group level and not the individual physician.

I think those three things, if they were enhanced, Ingenix might really -- you know, it would be really good advances, and we could do what you have been doing.

Actually, the Europeans have been tracking out sorts of care for years, and we still haven=t got a way of doing it. So, I think we need to get on the board.

CO-CHAIR ROSENTHAL: And I think much of tomorrow=s discussion is going to be around the general philosophic tenor of, can we by our actions help drive the next level of this? And I think that is going to be a lot of what tomorrow=s discussion is going to be about. So, it is going to be sort of open season for how could this be improved; how could this process be improved, et cetera, so that we tee this up for the people coming after us. But that will be all tomorrow, which Bruce is going to very ably direct us in.

DR. PETER: Hi. This is Doris.

Would there be a way to collect all the reasons why people voted no on the various measures, to give feedback to the measure developers? I know that is going to be part of the philosophical discussion tomorrow, but maybe more directed feedback might go beyond what we have already brought up.

MS. WILBON: Yes, we generally capture that in the meeting summaries and the report. So, we will definitely be capturing that. Thank you.

CO-CHAIR ROSENTHAL: All right. I think, with that, there is no break. We just keep moving.

But we are a little ahead of schedule. And so, we will move now to 1560, which is relative resource use for people with asthma from NCQA.

I think since this is the first NCQA measure that we have had today, perhaps we could just get a little precis of what this

measure is, and then we can open it up for discussion.

MR. HAMLIN: So, the NCQA measures are risk-adjusted, relative resource use for people with specific conditions. The methodology between the asthma and COPD measure is actually fairly similar, just a different chronic disease population.

They are reported out by service category, and NCQA currently only publicly reports information on entities that can provide a base population of 400 members or more. So, it is generally limited to health plans at this point in time. So that they are population-based measures for specific chronic disease populations.

CO-CHAIR ROSENTHAL: Can you just clarify, then, though, is that the level of attribution that is specified?

MR. HAMLIN: Yes, right now the level of attribution is open to anyone who has at least 400 people, 400 members who meet

their chronic disease definition. Right now, that has been only plans and a very small proportion of some large provider groups.

CO-CHAIR ROSENTHAL: Okay. Your earlier ones, if I recollect from the last meeting, specified group-level attribution or --

MR. HAMLIN: As long as they have a minimum sample size of 400 people and --

CO-CHAIR ROSENTHAL: Got it.

MR. HAMLIN: -- meet the definition, yes.

CO-CHAIR ROSENTHAL: Okay. All right. I think we have already voted that this is important, unless somebody feels it is not important.

But Paul?

DR. BARNETT: Perhaps he can also deal with those other two big issues that we had in the last set of measures, which was the pharmacy cost and whether the risk-adjustment method reflects any of the performance in the

period being evaluated.

MR. HAMLIN: So, as far as the pharmacy for asthma, pharmacy benefit is required for the measure because the quality measure that was reported alongside it is a pharmacy-based measure.

On the relative resource use side, the pharmacy is reported separately. So, if there is not a benefit offered, you will see a difference in the reporting result for the pharmacy, on the pharmacy side for the pharmacy utilization rate. However, since that is not rolled up in the total medical part of the RAU score, if you will, or the RAU result, you can see noticeable differences there. So, it is separate but equal, I guess is the way to put it.

For COPD, the pharmacy benefit is not required. So, that is probably where you will see the variability. But we do require the plans to provide, you know, to be accountable for obtaining the pharmacy data in

order to report the measures. And their scores are reflective of how well they do that.

I=m sorry, I just forgot --

CO-CHAIR ROSENTHAL: Did we get both questions?

DR. BARNETT: So, the second question is, does the risk adjustment reflect the performance in the current period, the procedures or outcomes in the current period?

MR. HAMLIN: Right. So, the risk adjustment was selected because it is the best method that we have found to inform for utilization, which is effectively what these resource use measures look at. It is dependent upon encounters, you know, because the weighting is based on number of identified diagnoses and/or -- so, people with multiple comorbidities, the comorbidity diagnoses you have, the increase in your risk score. So, you are weighted differently from those who have fewer. So, it is slightly affected by

that.

However, we found as a population-based approach it does a very good job of assigning members to specific risk cohorts based on the utilization for this total annual approach, again, because we are looking at every service that was delivered to these members.

DR. BARNETT: Just to follow up, so would specifically an asthma exacerbation or a COPD exacerbation during the measurement period get someone into a higher risk category?

MR. HAMLIN: Yes, it could.

DR. BARNETT: And how would that occur? By a different --

MR. HAMLIN: Well, there are 13 different risk cohorts. So, a patient is assigned to a risk cohort based on how many diagnoses, competing diagnoses, and other services they have encountered during the measurement timeframe. So, people with

multiple encounters for multiple exacerbations or multiple different diagnoses for different comorbidities would end up in a higher risk category, and therefore, it would be reported in that category. So, like I said, we have 13 risk categories right now.

So, someone who has just asthma and appears once perhaps for their regular visit during the measurement year would probably be in HCC Category 1; whereas, someone who has got multiple exacerbations might be in a 6 or 7 category because their frequency of service utilization is higher.

DR. BARNETT: So, based on the amount of utilization gets them into a higher category?

MR. HAMLIN: It is primarily the number of diagnoses that appears on their chart, which is generally affected by the number of times they have had some kind of encounter or some other service delivered.

DR. BARNETT: But if they had an

asthma exacerbation, aside from the fact that they have a chance to be coded for comorbidities, are there other ways in which their asthma exacerbation would contribute to a higher risk category?

MR. HAMLIN: Not specifically in every single case. So, theoretically, yes, an exacerbation would put them into a higher risk category, but, again, it sort of depends on what else on their chart for the measurement period.

DR. BARNETT: Maybe I wasn=t clear. I mean, other than the fact that they would have comorbidities coded from some other condition, co-occurring condition.

MR. HAMLIN: Right. So, an exacerbation could kick them up into a higher risk category, but 100 percent of the time I couldn=t say because it depends on individual patients, how many other comorbidities and other factors they have. It is a weighted risk adjustment. So, their weight score

increases as they have increasing number of services during the measurement period.

CO-CHAIR ROSENTHAL: And then, Ben, very quickly, and then we will move to the TAP report, remind us what the risk-adjusting methodology is that NCQA uses.

MR. HAMLIN: It is derived from the CMS HTC model.

CO-CHAIR ROSENTHAL: Okay.

MR. HAMLIN: So, it looks at, again, a series of diagnoses, and it ranks you and weights you based on age, gender, and number of other --

CO-CHAIR ROSENTHAL: And that is what you reported in the various others from the last meeting?

MR. HAMLIN: Yes.

CO-CHAIR ROSENTHAL: Right. Just clarifying.

MR. HAMLIN: It is the same across all of our e-measures.

CO-CHAIR ROSENTHAL: Okay.

Dolores?

MS. YANAGIHARA: So, does the number of times the diagnosis appears matter or is it just which diagnoses and the number of diagnoses?

MR. HAMLIN: It is number and types, yes. It is all factored in. Whether that takes you into another category, again, is dependent on how many and which category, you know, if you are going from a 6 to a 7 versus a 1 to a 2.

CO-CHAIR ROSENTHAL: And how many risk categories are there, then, when you end it? It is not a continuous variable?

MR. HAMLIN: No, there are 13 different discrete categories that you are assigned to.

CO-CHAIR ROSENTHAL: Yes. Okay.

DR. REDFEARN: Is that the standard way HTC works? My understanding was it doesn=t make any difference how many times you see a diagnosis; if it occurs once, it

triggers the grouper, and that generates the risk, and not the number of times --

MR. HAMLIN: We are not using groupers. We don=t use groupers for HTC.

DR. REDFEARN: For HTC?

MR. HAMLIN: Yes. So, the number of -- let me back up here. The diagnoses that are present during the measurement period for that patient will assign a specific weight to that patient. Competing diagnoses and other comorbidity diagnoses will, again, assign an additional weight. So, you basically, effectively, sum the weights of all the services rendered during that measurement timeframe, and that will be, once you have added your gender and age category weights, that will assign you to your specific risk cohort. So, there is a range for each risk category.

DR. REDFEARN: But the same diagnosis appearing more than once doesn=t make a difference? It has to be another

additional diagnosis?

MR. HAMLIN: Additional diagnoses.

DR. REDFEARN: Right.

MR. HAMLIN: So, yes, if you see asthma five times, you are not going to get into a different category. If you see asthma, COPD, and cardiovascular, right.

CO-CHAIR STEINWALD: Right. But if you have five different encounters, and in each one the diagnosis is asthma --

MR. HAMLIN: That won=t put you in a different risk category. It will put you in a higher utilization category.

CO-CHAIR STEINWALD: But not a different risk category?

MR. HAMLIN: Not a different risk category.

CO-CHAIR ROSENTHAL: Yes, I think you might have misspoken, because the first time you answered, you did say both the number of diagnoses and the number of frequency of their appearance.

MR. HAMLIN: The frequency only matters if you have different diagnoses --

CO-CHAIR ROSENTHAL: Got it. Okay.

MR. HAMLIN: -- not the same diagnosis. I=m sorry.

CO-CHAIR ROSENTHAL: All right. Barbara, do you want to clarify this?

DR. RUDOLPH: Well, the number of diagnoses is probably a proxy for the number of times you have had hospitalizations because they are much more likely to provide a much larger range of diagnostic codes than an individual practitioner.

So, someone who is hospitalized, has an exacerbation and is hospitalized, is going to end up with a lot more diagnoses than an individual who isn=t hospitalized.

MR. HAMLIN: Yes, but for chronic conditions, once you have been identified as having asthma, you will show up in the population. The number of other diagnoses

will put you in a higher risk category cohort, but the utilization component will be shown in the specific inpatient utilization scores for that particular --

CO-CHAIR ROSENTHAL: And then, this is one year all costs?

MR. HAMLIN: Any service during January 1st to December 31st for anyone identified with asthma. So, broken arms, scrapes, cuts, bruises, asthma exacerbations --

CO-CHAIR ROSENTHAL: You assume that is going to sort of spread itself out over the population?

MR. HAMLIN: Yes.

CO-CHAIR ROSENTHAL: And in asthma it probably does.

MR. HAMLIN: The idea is to get a picture of managing a person with this chronic condition, whether it is attributable specifically to the condition or not.

CO-CHAIR ROSENTHAL: And you don=t

think this one cross-reacts with some of the others like heart failure or COPD to a significant enough extent that episodes are going to get misattributed?

MR. HAMLIN: So, the specific exclusions attempt to minimize that, particularly with COPD. With heart failure, we recognize that there is some overlap for people with cardiovascular conditions, but we look at the specific population with asthma and then we look at the CV population separately, understanding there may be some overlap for that particular person, depending on where they end up.

CO-CHAIR ROSENTHAL: Okay. Can we, for the record, everybody believes that this is important, the same way we did the last time? Anybody who does not think it is important?

Thank you.

Now let=s move ahead with reliability and validity from the TAP.

So, Kurt, share your thoughts with us.

DR. ELWARD: Yes. Overall, the reliability is thought to be very good. It had very high ratings.

The results were repeatable.

One of the real challenges that, indeed, NCQA includes all costs. That means, if I had a little kid with asthma and he breaks his arm or he has a motor vehicle accident, that counts.

And overall, it was felt that it was very difficult to pull out, you know, decide which measure, which cost you would pull out, and that, for overall, patients with asthma, that those additional costs would not be very high, and over a large group of people would probably sort themselves out. But that was an issue. For asthma, we felt that those were rare enough that we could still accept that as a reliable criteria.

CO-CHAIR ROSENTHAL: Okay. Other

questions, comments, discussion for reliability?

DR. ELWARD: Oh, yes, I should say one thing.

CO-CHAIR ROSENTHAL: Yes.

DR. ELWARD: It was felt that a population of at least 400 members was needed for the methodology to be valid.

CO-CHAIR ROSENTHAL: Got it.

Other questions, comments, discussion?

(No response.)

All right. So, Ashlie, would you or Taroon tell us the TAP scores on overall reliability.

MS. WILBON: Overall? I=m sorry.

CO-CHAIR ROSENTHAL: Yes, I=m sorry, we are doing the subsegments and then overall.

MR. AMIN: Right. Okay. It is 2a1, well-defined, precise specifications, 9 high. 2a2, reliability testing, 8 high and 1

moderate.

CO-CHAIR ROSENTHAL: And then, overall?

MR. AMIN: Reliability overall, 8 high and 1 moderate.

CO-CHAIR ROSENTHAL: Okay. So, any further discussion?

(No response.)

I think we are ready to vote on this. This will be 1, high; 2, moderate; 3, low; 4, insufficient, and we are voting on 2a, overall reliability.

(Whereupon, a vote was taken.)

MS. FANTA: And for those of you on the phone, overall reliability.

Doris?

DR. PETER: High.

MS. FANTA: Hi.

(Laughter.)

High, moderate, low, or insufficient.

CO-CHAIR ROSENTHAL: She said

Ahigh@.

MS. FANTA: Blonde moment.

(Laughter.)

CO-CHAIR ROSENTHAL: Oh, that was cute. I missed that completely. AOh, hi.@

MS. FANTA: Yes, exactly.

CO-CHAIR ROSENTHAL: AHow are you?@

(Laughter.)

I think we are all getting a little punchy.

MS. FANTA: Ethan?

DR. HALM: High.

MS. FANTA: Okay. So, we have 12 high and 3 moderate.

CO-CHAIR ROSENTHAL: All right. Great.

Now let=s move to the next part, which is validity.

And, Kurt, the TAP view on this?

DR. ELWARD: The face validity. Overall, they had high scores. The face

validity was clear, but the categorizations based on age weren=t very clear.

There was in-depth discussion regarding the measure exclusions. NCQA explained that they are used in the risk adjustment -- I=m sorry. Wait a minute. I think I am ahead of myself here. Yes.

Overall, the scores on validity were high. I=ll put it that way.

CO-CHAIR ROSENTHAL: Okay. Open for discussion. Barbara?

DR. RUDOLPH: I was just wondering about the pharmacy cost. Do you have some way of knowing whether or not, even though they might have a pharmacy benefit, whether or not a PBM might have withheld the cost information? Or is there a way to exclude cases like that?

MR. HAMLIN: There=s no way to exclude cases like that currently. We do have a way to determine that, but it does require going back to both the auditor and the

submitting organization to determine if that was one of the factors that affected their pharmacy score. It is not directly part of the reporting strategy.

So, we do see the different rates within the pharmacy scores. But, again, looking at fluctuation of those scores in comparison to another plan that is determined to be in the peer group, the only way you can tell the significant difference is because of some kind of design issue. We would be going back through the audit process to determine what factors might have informed that specific result.

CO-CHAIR ROSENTHAL: Is there a way to game the encounter submission? That is the thrust of your question, right?

DR. RUDOLPH: The thrust was just that -- actually, Jack pointed this out to me -- that in the Ingenix measures, and probably in this too, you know, it is required to have a pharmacy benefit. But, then, if the

pharmacy claims are handled through a PBM, you don=t actually get the cost back unless they go back and the plan actually requests very specific costs from the PBM. So, it will show up --

CO-CHAIR ROSENTHAL: But how is that different than what Jack has been asserting all along?

DR. RUDOLPH: It=s not different.

CO-CHAIR ROSENTHAL: Oh, okay. I=m sorry.

DR. RUDOLPH: I just want to make it clear that it is not any different than --

CO-CHAIR ROSENTHAL: Oh, I=m sorry.

DR. RUDOLPH: -- what the Ingenix situation was.

CO-CHAIR ROSENTHAL: Yes, yes. Okay.

MR. HAMLIN: So, we are not actually looking at actual cost for the pharmacy. So, the pharmacies are all priced

in a standardized pricing, like our other services are as well. So, they don=t need the actual price of the pharmacy that they are paid. But as long as they can track the code for the pharmacy that was delivered, it will be included.

CO-CHAIR ROSENTHAL: Okay. Jack, do you want to --

MR. BOWHAN: Well, that would be the point about the PBM. If you are not getting the claim, you don=t have whatever cost of using it --

MR. HAMLIN: We say the plans are responsible for obtaining that data to report the measure. It is up to them to determine how much they want affect their score and how much they want to push the PBMs to give them the data they need.

CO-CHAIR ROSENTHAL: But, again, just to clarify, I=m back to this. If a particular plan or entity simply does not get the pharmacy benefit because it is completely

carved out and the pharmacy benefit isn=t available to them, that wouldn=t be scored, then, correct?

MR. HAMLIN: Well, their score would be affected probably for that one entity that they could deny the data. You would see a difference in the pharmacy ratio.

CO-CHAIR ROSENTHAL: All right. So, yours does not handle it the way Ingenix did, which was basically to exclude the pharmacy cost for any entity that doesn=t --

MR. HAMLIN: No.

CO-CHAIR ROSENTHAL: Well, wait. I am just trying to clarify. I could be wrong.

MR. BOWHAN: I don=t think Ingenix automatically excludes it. They suggest that whoever is running the report do that. But on the normal, standard reports that they have coming out, it is not necessarily excluded, and separating out patients who don=t have a pharmacy benefit from those who do, to my

knowledge.

CO-CHAIR ROSENTHAL: Well, we didn=t approve the Ingenix one anyway. But my understanding of what I understood the answer to Jack=s question around the Ingenix was is that, if you were an entity that didn=t have pharmacy benefits, you didn=t get scored in comparison to an entity that did.

DR. RUDOLPH: You might have pharmacy benefits, but they are run through a PBM. So, yes, they would be included, but they may not have the information from the PBM to actually incorporate.

CO-CHAIR ROSENTHAL: Okay. All right, I got it. But here it is moot because this is standardized pricing, right?

MR. BOWHAN: Not if you don=t the claim. You have to get the claim to generate the standardized pricing.

CO-CHAIR ROSENTHAL: Right, but that is true of anything. So, that is back to my question. I mean, if an entity is going to

game it by excluding claims, or whatever, it is only to their own detriment.

MR. HAMLIN: The audit process generally removes any kind of gaming in the withholding of claim information to ensure that. So, all these data are audited prior to being submitted and verified by a certified auditor before being submitted to NCQA. So, a lot of that, we try to hit that before it comes to us.

CO-CHAIR ROSENTHAL: So, if it is a PBM and they choose not to give the claim data to the plan, period, it could affect it. But what health plan is going to be in that setting where they are not going to get, insist on getting the full claim data? And then, you are left with the question, well, is somebody gaming the claims data? And the answer is there is an audit process, right?

MR. HAMLIN: Yes. It would pick that up if it was a major issue.

DR. NEEDLEMAN: Ben, I have got to

admit that this conversation has totally confused me.

(Laughter.)

And has to do with, okay, you have got an audit process. First of all, my understanding is what you said was the plan is obligated to get at least the pharmacy claims as a file of here are prescriptions for our patients. So, you know what was prescribed. And then, you have got a standardized pricing module for imputing cost to that. Okay.

But, then, I heard you say something about where your score is, which implies that somebody can not be getting either some of that data or all of that data, but still be in your system. And that is what confused me.

MR. HAMLIN: So, we look at all, for the RAU, we look at all pharmacy dispensed. So, any claim for a dispensed pharmacy would end up in the RAU score.

We require for the asthma

measurement, in particular, that they have a pharmacy benefit. Whether they have the complete claims for all of their members is up to the plan to determine that they have comprehensive claims, and there is an auditor that has to go in and verify that they, in fact, have complete datasets before they submit the measure to NCQA.

So, there is a way that potentially incomplete data could affect their calculated score and their result, but that is generally minimized by the auditors going in and ensuring that all data fields that are required to report the measure are complete, and that they are being submitted properly and calculated properly for NCQA.

DR. NEEDLEMAN: So, just again, in contrast to what we were hearing, if the University of California carves out its pharmacy benefits to CVS, and they do not collect the pharmacy claims to run through the Ingenix grouper, we wind up in the category of

no pharmacy, in the stratification of no pharmacy. But they could not submit their data to NCQA because there is no pharmacy benefit database there.

MR. HAMLIN: They would probably submit the pharmacy index as an NA. You know, so they would not be able to report that because they would not have complete data for the pharmacy.

DR. NEEDLEMAN: And so, what happens in that case?

MR. HAMLIN: They are still able to report RAU because, again, we have the total medical, we have the quality, and we have the pharmacy, which are separate components of it. So, they are allowed to have a certain number of missing components, you know, and still be able to submit the measures to us.

But, again, we hold the plans accountable for ensuring that they have the complete data that is submitted to us in order

to report the measure.

DR. NEEDLEMAN: So, measured against either the number of plans that you are providing data for or the number of groups that those plans are representing, I am not quite sure what level we are talking about here, what proportion do not have pharmacy data? What is the proportion that are pharmacy data NA?

MR. HAMLIN: I don=t have that information at my fingertips. But we right now have 374 commercial plans and 190 Medicaid and 103 Medicare plans that are reporting complete data. So, there=s a number of plans above and beyond that that are not able to report, probably due to some issues either in the pharmacy or on the other medical side. So, they don=t end up in the final calculation because either they do not have the required benefit or they do not have the required information to report the measure.

CO-CHAIR STEINWALD: This issue

has come up before, the difference between what happens in practice and the measure that you are seeking NQF endorsement for. That measure includes pharmacy benefits, right?

MR. HAMLIN: Right.

CO-CHAIR STEINWALD: Okay.

MR. HAMLIN: And the measure specification details exactly what is required to report the measure.

CO-CHAIR ROSENTHAL: And I think a little bit of the disconnect, Ben, is that the example that Jack used was the University of California, and let=s assume it was an ACO of some ilk, but probably would not have access to the pharmacy benefit programs for multiple health plans, would not be able to submit, or if it did submit, would certainly not have pharmacy benefits. And yet, the real-world experience of your organization is, and why it requires 400 individuals is, it is health plans, and health plans virtually almost every time have access to the pharmacy encounters.

***MR. HAMLIN:** I mean, again, they would be able to submit whatever data they would wish to us, based on the specifications, but they would not end up in any of the reporting products because they were missing a major component of the measure spec.*

***CO-CHAIR ROSENTHAL:** But most of the health plans do have or many --*

***MR. HAMLIN:** Right now, about a little over two-thirds of the plans that report, all the plans that report to us, report RAU successfully. So, the number is increasing. It went up 8 percent this year from last year.*

***CO-CHAIR ROSENTHAL:** Okay.*

***MR. HAMLIN:** So, increasingly, we require them to get the data, and they are going out and finding it.*

***CO-CHAIR ROSENTHAL:** Got it. Okay. But, again, a provider entity would not have the leverage in most instances --*

***MR. HAMLIN:** There=s a whole other*

series of issues there in that, yes.

CO-CHAIR ROSENTHAL: There=s a whole other series of issues, right.

Okay. We are on 2b, overall validity.

DR. REDFEARN: I have another --

CO-CHAIR ROSENTHAL: David?

DR. REDFEARN: -- issue to raise.

CO-CHAIR ROSENTHAL: Absolutely.

DR. REDFEARN: One of the things that struck me in going through this is the fact that you use indirect standardization when you do the risk adjustment. Why did you choose indirect standardization?

My concern is, when you are trying to reweight a small organization=s distribution of whatever you are comparing them on based against the overall averages, you may be weighting relatively rare occurrences for that organization pretty substantially and underweight other things that they are doing.

I mean, I even misinterpreted this the first time around. I thought it was direct standardization because everybody has been using that, but you are using indirect standardization. And I wondered what the logic of that, why that was done that way.

MR. HAMLIN: I wasn=t involved in the development phase. So, the ultimate decisions were -- but my understanding is that, during the testing when they were trying to determine what the most equitable and reliable standard for the measure specification, that was sort of what they landed on for their calculation.

I mean, I would agree there probably are some specific smaller plans that may be more greatly affected in this, but, again, overall, for the national plan reporting of the 850 or 900 plans that report to NCQA, I think those probably are minimized. It is not the perfect approach, but it is the best of what works for plan-to-plan

comparisons at this time.

DR. NEEDLEMAN: *Do you want to comment on direct standardization versus indirect and what you see as the strengths and weaknesses of each? Because you, clearly, have thought about this.*

DR. REDFEARN: *Well, when you do direct standardization, basically, you adjust the norm to match the distribution for the entity that you are comparing it to. So, that sort of gives the advantage to that organization to say, AI=m going to evaluate you based on your particular mix of services or risks,@ or something like that. That is the way I have always done it, and that is the way we do our provider profiling and stuff like that.*

They are doing the reverse. They are saying we have a distribution that we have derived from all of our aggregate data put together, and we are going to use that distribution for every group we are comparing,

no matter what their distribution is.

And you can think of really absurd cases in which you are looking at a group that has a mix that is very different, very atypical. And in that case, you are going to heavily weight things that they just don=t do very much about. And that means you are taking a very small number and you are projecting it out to do part of your evaluation, which just makes me really nervous.

There are arguments in both areas in terms of the provider profiling world. If you know anything about Doug Cave and his approach, Doug recommends indirect standardization for everything he does in provider profiling because he says you do specialty-specific comparisons, and what should a rheumatologist be doing? A rheumatologist should do what rheumatologists do on average, and that is how I am going to compare everybody that is says they are a

rheumatologist.

We don=t do it that way. We think that leads to some potential misunderstanding. But it is a legitimate argument. That is the distinction.

It is just the odd thing here is, I mean, this is the first situation we have seen in which it is indirect standardization.

MR. HAMLIN: And I think, partly, that may also be due to the fact that our smallest reporting entity right now is an HHS region, which is actually fairly large. You know, this is not part of the spec, but we are looking at increasing the specificity of the regional component of the RAU measures. So, we would love to get down to HRR or HSA, if we could, but to be addressing the market variation.

But, right now, we have to calculate a national and an HHS region, which is a pretty big slice in which a lot of variation occurs. And I think that is

probably why it was the best approach for the current approach.

Should we get more granular in the future, I think we may revisit that issue. But, right now, since the largest entity is the HHS region, which encompasses several states and many different markets, there is just too much variability, I think, within that region to apply a direct standardization approach, I wouldn=t say easily, but sort of reliably, that would apply to a West Coast region versus an East Coast region or something along those lines.

***DR. REDFEARN:** When you have huge entities, it probably doesn=t make any difference anyway.*

***MR. HAMLIN:** It may or may not.*

***CO-CHAIR ROSENTHAL:** Other questions on overall validity?*

(No response.)

All right. Then, Ashlie, if you would give us the TAP scores? And then, we

will call the question.

DR. RUDOLPH: Could I ask one more thing?

CO-CHAIR ROSENTHAL: Yes, ma=am.

DR. RUDOLPH: It just took me a minute.

In the standardization, do you separate out the commercial plans from the Medicaid and Medicare?

MR. HAMLIN: Yes, each product line is calculated completely separately from each other. So, your peers are only being compared to peers.

DR. RUDOLPH: Okay.

MS. WILBON: All right.

DR. NEEDLEMAN: I=m sorry. I=m looking at the TAP report summary report. And what level is this reported at? Because the TAP report says it goes down to the clinician level. Is that accurate?

MR. HAMLIN: No.

DR. NEEDLEMAN: Okay. What level

of reporting are we talking about here?

MR. HAMLIN: We use it for health plan reporting. But, again, it could be used for anything with a population of at least 400 members.

CO-CHAIR ROSENTHAL: Can we have the TAP?

MS. WILBON: So, 2b1, specifications are consistent with the resource use or cost problem, 6 high, 3 moderate. Validity testing, 6 high, 3 moderate. Exclusions, 6 high, 3 moderate. Risk adjustment, 7 high, 2 moderate. Identification of statistically-significant and meaningful differences, 8 high, 1 moderate. Overall validity is 5 high, 4 moderate.

CO-CHAIR ROSENTHAL: Okay. Any further discussion?

(No response.)

It is amazing, just when you think you have discussed every possible aspect of

this, there is, in fact, some point of this thing that is actually tremendously helpful for the group in our education.

Helen, just when we actually learn something, we become useless.

(Laughter.)

All right. So, I think we will call the question.

DR. BURSTIN: I must admit, I don=t think I picked up before on this point, but it is health plans or an AN@ greater than 400. I mean that is actually pretty significant.

CO-CHAIR ROSENTHAL: No, it=s good.

DR. BURSTIN: Yes.

CO-CHAIR ROSENTHAL: And it could get down to an individual provider who is unbelievably busy seeing asthmatics. But that, I think, in the world we live in doesn=t exist.

DR. BURSTIN: But it is

potentially very applicable to the sort of emerging models of --

CO-CHAIR ROSENTHAL: Yes, emerging models.

MR. HAMLIN: Asthma is actually one of the conditions that is most affected by this because there are actually a number of plans that cannot meet that minimum sample size requirement. So, that is the one where most plans get limited --

CO-CHAIR ROSENTHAL: Except for the pharmacy benefit problem, which, again, most of the ACOs don=t have access to their pharmacy benefits.

And again, if we are ever really going to have integrated delivery, we have got to have really integrated data and that the people know what the heck is going on. But those are all editorial comments.

(Laughter.)

Let=s vote. So, this is 1, high; 2, moderate; 3, low; 4, insufficient.

(Whereupon, a vote was taken.)

MS. FANTA: And on the phone, overall validity.

Doris?

DR. PETER: High.

MS. FANTA: Okay. And Ethan? Ethan, are you still there?

(No response.)

Okay. So, we have 4 high, 9 moderate, and 1 low.

CO-CHAIR ROSENTHAL: All right. And now we get to vote overall scientific acceptability. Our options are more limited again. So, this is yes or no; 1, yes; 2, no.

I am not going to ask for any more conversation because, when I do, I get it.

(Laughter.)

Which until about 30 seconds ago was a very good thing.

So, 1, yes; 2, no.

Sarah, tell us when you are ready.

(Whereupon, a vote was taken.)

MS. FANTA: And Doris, overall scientific acceptability?

DR. PETER: Yes.

MS. FANTA: Okay. So, we have 12 yes and 2 no.

CO-CHAIR ROSENTHAL: Great. Let=s move to usability.

Kurt, I think we will move right to the TAP discussion.

DR. ELWARD: Yes. Yes, I think, overall, there was high, generally high levels of votes for usability. There was a concern about how smaller groups would implement that. I think Ben has addressed that. Smaller entities would have a problem doing this. But, otherwise, the majority of people who would use it would have been able to do it well.

CO-CHAIR ROSENTHAL: Okay. Discussion about usability?

Paul?

DR. BARNETT: I was just

wondering, so the process requires that the plan actually turn over data to NCQA to actually run it and do it?

MR. HAMLIN: Yes. Plans provide aggregate data on PMPM to NCQA for all the members who meet the criteria for each service category. So, not member-level information.

CO-CHAIR ROSENTHAL: I guess, is, Paul, your question, though, could some other entity take this measure and apply it to some group that had 400 members, knowing how to use the risk-adjusting methodology, et cetera, et cetera, et cetera? In other words, does it specify that it is only NCQA that can apply the measure?

MR. HAMLIN: No. All our methodology is transparent. We put it on the website. So, any entity that wanted to do the same thing could do the same thing. It is a distributed model, though.

So, the number of plans that report the measure to NCQA allows us to

calculate better expecteds for each plan average for each of the service categories. So, it helps to be NCQA, but anyone can do it.

DR. BARNETT: And I am not sure whether this is usability or feasibility, but they have to be an NCQA subscriber, member, or something like that, to --

MR. HAMLIN: You do not have to be an accredited plan to submit data to NCQA. You are able to submit, and we will return you a calculated IDSS report, whether or not you are accredited. It cost you a little bit to do it, but there is no requirement for accreditation to submit the data to get the report back -- a little bit less than it does for accreditation, I think probably.

CO-CHAIR ROSENTHAL: Or you could do it yourself, but you would have very little to compare it to.

MR. HAMLIN: Right.

CO-CHAIR ROSENTHAL: Right. And the observeds-to-expecteds would be hard to

calculate.

MR. HAMLIN: Yes. Well, you would have the observeds, just not a very good expected.

CO-CHAIR ROSENTHAL: You wouldn=t have a very good expected, right.

MR. HAMLIN: Yes.

CO-CHAIR ROSENTHAL: Right. Okay.

Any further question/discussion on usability?

(No response.)

All right. Hearing none, let=s hear the TAP -- we didn=t do this. I=m losing my mind. What was the TAP vote on this?

MS. WILBON: On 3a, whether or not the measure performance results are publicly reported, 8 high; 1 moderate. 3b, whether or not the measure is meaningful, 6 high; 3 moderate. And whether or not the measure is transparent is 8 high; 1 moderate.

CO-CHAIR ROSENTHAL: All right. And so, our vote is on overall usability, and

it is 1, high; 2, moderate; 3, low; 4, insufficient. And let=s vote.

(Whereupon, a vote was taken.)

MS. FANTA: And Doris, on usability?

DR. PETER: High.

MS. FANTA: Okay. So, we have 9 high and 5 moderate.

CO-CHAIR ROSENTHAL: All right. And last, then, we have feasibility.

And, Kurt, the TAP view of this?

DR. ELWARD: Yes, just to say Ashlie some time, it was sort of 9, 7, 8. They were all very high levels.

And the data elements are available electronically.

CO-CHAIR ROSENTHAL: Right, it is coded information.

DR. ELWARD: Yes, it is coded information.

CO-CHAIR ROSENTHAL: It is claims with both the positives and the limitations of

claim use.

DR. ELWARD: And NCQA does a good job of recognizing where there are challenges with data inaccuracy.

CO-CHAIR ROSENTHAL: Okay. Open for discussion.

(No response.)

All right. Hearing none, I will take that that we are ready to vote on this. One, high; 2, moderate; 3, low; 4, insufficient.

(Whereupon, a vote was taken.)

MS. FANTA: Okay. And Doris, your vote on overall feasibility?

DR. PETER: High.

MS. FANTA: Okay. So, we have 10 high, 4 moderate.

CO-CHAIR ROSENTHAL: All right. And now, we are left to vote on overall recommendation for endorsement.

I don=t think we get a TAP vote on this, do we? No, we just have to do this

ourselves. Okay.

(Laughter.)

I=m getting tired.

Okay. So, this is easy. It=s yes or no or abstain. And now, we are voting on recommendation for endorsement or a no vote is against endorsement.

So, with no further discussion, let=s vote.

(Whereupon, a vote was taken.)

Oh, wait, I voted wrong. What do I have to do? Oh, I can change it? Okay.

MS. FANTA: And Doris, your vote on the overall recommendation, yes or no?

DR. PETER: Abstain.

MS. FANTA: Okay. So, we have 13 yes and 1 abstention.

CO-CHAIR ROSENTHAL: All right. So, that concludes the discussion on 1560.

Now let us take up 1561, which is relative resource use for people with COPD from NCQA.

And again, I think in sort of the interest of time and people=s sanity, if we could focus on what aspects are different from asthma without necessarily going through every element of the measure, we might be just a smidge more efficient.

Let=s go. Let=s push through.

All who believe that this important?

Anybody believe that it is not important?

Okay. So, importance is settled.

Ben, do you want to give us the quick version of COPD? And again, focus on how it is similar or different to the asthma measure.

MR. HAMLIN: It is different because it applies to COPD and not asthma, the same service categories, the same risk-adjustment approach, the same standard pricing tables.

CO-CHAIR ROSENTHAL: Okay.

MR. HAMLIN: It=s COPD, not asthma.

CO-CHAIR ROSENTHAL: Well, that=s what I wanted to get. It is a different diagnosis, but, otherwise, the methodology is the same?

MR. HAMLIN: Yes, it uses different diagnosis codes from ICD-9 to identify people with COPD, and pretty much everything else is the same.

CO-CHAIR ROSENTHAL: All right.

So, Kurt, let=s do reliability.

DR. ELWARD: Yes. They use some more measures. The populations are a little bit different in terms of it is a little harder to do fee-for-service for the general eligible population of Medicare. But, overall, our ratings for reliability were high.

CO-CHAIR ROSENTHAL: Okay. Open for discussion.

(No response.)

I have one question. I am not sure whether reliability is the place to ask it, but it gets to the business about intercurrent diagnoses with COPD seem to be much more likely than they were with asthma. And how is that managed in your world about this? So, you have got people with heart failure and potentially multiple other chronic diseases, particularly in the elderly.

MR. HAMLIN: Right. So, for COPD, there are fewer exclusions, clinical exclusions. For the asthma population, we try to exclude the emphysema/the COPD from that population, so they will end up in the COPD RAU measure.

For other diagnoses like heart failure, they will be risk-adjusted in a different category than somebody who does not have that comorbidity, but that would be where they would be differentiated, is in the risk adjustment.

CO-CHAIR ROSENTHAL: And then, an

individual patient could end up both in a COPD episode grouper, as it were, and also a CHF measure, and be risk-adjusted appropriately for both diagnoses in both populations?

MR. HAMLIN: Right. You assign the diagnoses and you take the highest ranked one when you do the HAC risk adjustments. So, yes, they are all factored; they are all taken into consideration. So, yes, depending on however many of those they have, they will be adjusted appropriately, depending on how many diagnoses that they have.

CO-CHAIR ROSENTHAL: But could a patient in a health plan end up in two different diagnostic groups? Or is it literally only the primary diagnosis?

MR. HAMLIN: For risk adjustment, you take all diagnoses, not just primary.

CO-CHAIR ROSENTHAL: All right. No, I am talking about, could a patient -- I=m in a health plan. I=m in Blue Cross of Ohio, and I have COPD and heart failure. And there

is a heart failure metric cost of care and there is a COPD metric cost of care.

MR. HAMLIN: Right.

CO-CHAIR ROSENTHAL: Can I end up in both of those groups?

MR. HAMLIN: If we had a heart failure one, then yes.

CO-CHAIR ROSENTHAL: Yes, if you had a heart failure one. Okay. And I am not saying there is anything wrong with that.

MR. HAMLIN: Yes.

CO-CHAIR ROSENTHAL: That is just for clarification purposes.

Sorry.

DR. ELWARD: Yes, that is one of the challenges. In some ways, you have one person splitters --

CO-CHAIR ROSENTHAL: Yes.

DR. ELWARD: -- and NCQA is a lumper, with all due respect.

(Laughter.)

But there is so much variability

that that was a real concern, but, again, on the one hand, asthma has very few comorbidities. So, we think it is going to sort out.

CO-CHAIR ROSENTHAL: Right.

DR. ELWARD: But, at the other end of the spectrum, the thought was that COPD folks overall have so many comorbidities that that may sort itself out. The question is, is that really accurate?

CO-CHAIR ROSENTHAL: Yes. Well, if they risk-adjust it, it is no problem, and it sounds like they risk-adjust it.

DR. ELWARD: The problem was that two things really drove our recommendations. One is that they do risk-adjust, and fairly well, and second, that the process was transparent. So, we could understand how they did that.

CO-CHAIR ROSENTHAL: And you could have an individual who ends up in both sets for cost, and yet, they both get risk-adjusted

appropriately. Okay.

Other questions on overall reliability?

Jack?

DR. NEEDLEMAN: Yes. Since we are talking about risk adjustment, and I am never sure whether it is reliability or validity, you had mentioned the broken arm. I think of the person who gets hit by the bus. You know, how are things like getting hit by a bus or being diagnosed with cancer, but, in particular, those acute things, are those built into your risk-adjustment model? Or are you just relying upon we=ve got 400 people at least and it is going to average out over --

MR. HAMLIN: So, we have, as part of the tables that we post for our risk-adjustment methodology, we have, I think, 187 different clinical conditions that are identified that you have to look for for the risk adjustment. So, if they are on that list, then, yes, they are included in the

risk-adjustment method.

I haven=t looked at hit by the bus on the table yet, but I=m sure fractures and other things would be included as part of those.

You know, you get assigned for some other service, some other encounter that you have had, some other diagnosis of AX@, if you will.

CO-CHAIR ROSENTHAL: Are there any exclusions?

MR. HAMLIN: Well, there are mandatory exclusions for all of RE measures, which are HIV, active cancer, ESRD.

CO-CHAIR ROSENTHAL: That=s right, we dealt with this the last time.

MR. HAMLIN: So, those are automatically excluded from the measurement altogether. So, they are sort of the high-cost conditions where a few patients could really skew the results for one plan.

CO-CHAIR ROSENTHAL: Even with 400

members.

MR. HAMLIN: Even with 400 members. Transplantation is the other one. So, high-cost conditions that are --

DR. NEEDLEMAN: And some of the other things, like these acute --

MR. HAMLIN: Right.

DR. NEEDLEMAN: -- acute high-expense incidences --

MR. HAMLIN: Right.

DR. NEEDLEMAN: -- are sort of being picked up by your risk adjustment?

MR. HAMLIN: Some of those will be picked up by risk adjustment. Some of those will show up, if there are a number of those, they will show up in the specific service categories. So, we look at acute inpatient. We look at ED discharges, and those kinds of things, as part of the measure specification. So, you will see them.

Most of them, I believe, will be captured by risk adjustment for sort of the

187 standard clinical identifications, but, also, again, by reporting out by specific service category, acute inpatient/non-acute inpatient, those kinds of service categories. You will see if you have a lot of people who are playing in traffic for that one year who happen to have COPD, that will show up in their specific results.

CO-CHAIR ROSENTHAL: Okay. Let=s look at the TAP results. I think it is becoming clear how this is going, but let=s do that. And then, we will vote on overall reliability.

MR. AMIN: 2a1, well-defined, precise specifications, 9 yes -- or 9 high. And 2a2, reliability testing, 8 high; 1 moderate.

CO-CHAIR ROSENTHAL: Okay. And did they vote overall reliability?

MR. AMIN: Yes. Seven high, 2 moderate.

CO-CHAIR ROSENTHAL: Okay. All

right. So, then, I think we are prepared to vote overall reliability, No. 2a, which for us, again, is 1, high; 2, moderate; 3, low; 4, insufficient.

And so, if we are prepared, let=s vote on this.

(Whereupon, a vote was taken.)

MS. FANTA: And Doris, your vote on overall reliability?

DR. PETER: High.

MS. FANTA: Okay. So, we have 11 high and 3 moderate.

CO-CHAIR ROSENTHAL: Okay. Let=s do validity now.

Kurt? Put your microphone on.

DR. ELWARD: Yes. Again, it goes back to what Jack was talking about, multiple comorbidities. So, I think we have had that discussion already.

In general, the ratings were high because the treatment of outliers were tagged appropriately. You know, the biggest driver

is severity of disease, but it appears that they are risk-adjusting as well as we could expect.

CO-CHAIR ROSENTHAL: Okay. Open for discussion.

(No response.)

Somebody surprise me with an issue that we have not discussed. Not possible. All right, don=t test it. Don=t push our luck.

(Laughter.)

I tried.

So, I think we are ready to vote. So, let=s go through the TAP scores there.

MR. AMIN: On 2b1, specifications consistent with the resource use and cost problem, 8 high; 1 moderate. 2b2, validity testing, 8 high -- or 6 high; 3 moderate. 2b3, exclusions, 4 high; 5 moderate. 2b4, risk adjustment, 6 high; 3 moderate. 2b5, identification of statistically-significant and meaningful differences, 5 high and 4

moderate.

CO-CHAIR ROSENTHAL: All right.

MR. AMIN: And overall was 4 high and 5 moderate.

CO-CHAIR ROSENTHAL: Okay. So, we will be voting overall validity, 2b, and our votes are 1, high; 2, moderate; 3, low; 4, insufficient.

And let=s vote.

(Whereupon, a vote was taken.)

MS. FANTA: And Doris, your vote on overall validity?

DR. PETER: High.

MS. FANTA: Okay. So, we have 4 high and 10 moderate.

CO-CHAIR ROSENTHAL: All right. Now we vote overall scientific acceptability, and this is yes or no; 1, yes; 2, no.

And let=s vote.

(Whereupon, a vote was taken.)

MS. FANTA: And Doris, overall scientific acceptability?

DR. PETER: Yes.

MS. FANTA: Okay. So, we have 13 yes and 1 no.

PARTICIPANT: No, I pushed the wrong thing.

MS. FANTA: So, we have 14 yes.

CO-CHAIR ROSENTHAL: All right. So, now usability.

Kurt?

DR. ELWARD: Generally, the same thing. One of the things that the TAP did appreciate was that NCQA does extensive audits on their material on a regular basis, and you can deconstruct the measure to facilitate transparency, which we thought was very important.

CO-CHAIR ROSENTHAL: Okay.

DR. ELWARD: So, it is not only user-friendly in terms of use, but also interpretability and being able to be deconstructed.

CO-CHAIR ROSENTHAL: Okay. And

just for completeness sake, the only lever on that is that, even though you could do this on your own without going through NCQA, it would not be completely trivial. Right. Okay.

DR. RUDOLPH: One question.

CO-CHAIR ROSENTHAL: Yes, ma=am?

DR. RUDOLPH: How would this data be used by non-plan personnel, by a provider, for example? How would the results be used?

MR. HAMLIN: So, what we have seen so far is that, because this gives you a snapshot of utilization for these chronic disease conditions, we found that this allows participating healthcare services to have much more information when they go into negotiations for their next annual purchasing time. So, they can look at the premium. They can look at their relative resource use. They can look at their quality score. And they can ask some harder questions about, well, why are you here versus that other plan is over here.

So, it really is that the

purchasers we have found have been really interested in this. The plans also have been interested in going back, applying the same methodology, plugging in their own actual prices or their allowed prices, or whatever they choose to do, to identify opportunities where they might have effect. You know, so much effort in one of these service categories might have a much greater effect than a greater effort in another category, just depending on what the utilization is. And we offer programs that help them do those calculations to try to make the results more meaningful.

DR. RUDOLPH: Have you sort of looked at longitudinally whether changes have resulted?

MR. HAMLIN: We are trying to figure out a way to do that right now. The level of data that we get, and because we do this calculation every year, we can=t trend the data directly. But we are looking at ways

now, as we automate more of the data collection.

There are about 5,000 data elements per measure per plan that come in. So, we are trying to (a) reduce the burden, but we are automating a lot of this, so we can try to hold the pricing and other things constant over multiple years, as we get multiple years of data, and do calculations that way.

We just haven=t done that yet because of the level of computing power. We just haven=t had the ability to do that yet, but we are hoping to starting this year, moving forward. So, in three years= time, we could go back and recalculate things, holding a bunch of things constant, and show trendability. But that is a computer-level, a server-level issue up to this point. It is a lot of power that is required.

CO-CHAIR ROSENTHAL: All right. So, we are going to vote on usability.

Jack, please.

DR. NEEDLEMAN: This probably is a comment that is more suited for tomorrow=s discussion, and it isn=t going to affect my vote. But I just think, as we go through all these measures of resource use, it is important to keep in mind that, ultimately, what we have got and what we are analyzing are only resources that are billed for.

Any service that a health plan or a physician group or an employer, for that matter, is providing to support particularly people=s efforts to manage their own chronic illnesses, are simply not captured as resources that we are measuring and will not be taken into account in understanding differences in performance of different plans or employers or provider groups in delivering effective care.

And that=s okay. That=s where we are in terms of what data we have available for this. But it is just important to keep

that in the back of our minds as we go through labeling these the resources that are being consumed in delivering care.

MR. HAMLIN: While we don=t measure them directly, we actually do feel that programs like wellness and DM programs do have an effect on the results. So, again, it is not a direct measurement, but we do feel that, because we are reporting these out by specific service categories, you might see a shift from inpatient to more outpatient E&M if you have a really good wellness program that is identifying risks in the population.

So, we do say that. We say we feel that these programs, while not directly measured, will affect your results, and therefore, we support the continued use of good wellness programs and risk identification in your population, and screening.

CO-CHAIR ROSENTHAL: Joe, did you have a comment that you wanted to make on that?

DR. STEPHANSKY: ***What we are seeing in Michigan from some of the plans are specific proprietary, essentially, CPT codes covering some care coordination issues. And we are seeing a lot more of that very quickly as the patient-centered medical home comes.***

So, someplace along the way, if we don=t have anything to map those to, and we are only mapping them back to codes that can be used on existing bills, we are going to run into a problem. I think there is an opportunity here, but I don=t know how to make use of it.

MR. HAMLIN: ***We price services that we can price reliably over a large scale. So, we have problems with a few services that are very proprietary or very unique to certain areas.***

We are working right now on the quality side of this to look at programs as we respecify measures for CMS for EMRs, ambulatory, to meet meaningful use in all

these things. We are hoping that some of these care coordination issues might be rolled into the quality side at first, until we can somehow figure out how to get them on the resource use and how to directly measure those.

Again, our standardized price schedule is basically Medicare fee-schedule-based with some adjustments for commercial utilization.

But you=re right, there are some really great programs that we just can=t measure right now. We want to; we just can=t.

CO-CHAIR ROSENTHAL: Barbara?

DR. RUDOLPH: Yes, I think, especially for like COPD, transfers from hospitals to institutions, long-term care units, et cetera, those things are not being included at all, as Jack mentioned. So, we do have to sort of think about how will we make that integration between those very costly services that aren=t being measured,

particularly if there is sort of provider failure. You know, those folks, the train wrecks are more likely to go to long-term acute care facilities or long-term care hospitals. And somehow, we have got to get at those kind of costs.

MR. HAMLIN: Yes, it is one of the ironies in our HEDIS quality measurement side where, for COPD in particular, we have assessment and we have management of patients, and none of the care management in between that really is very important to managing COPD members.

And we are hoping, again, with ambulatory-based EMRs that are very granular, with the measurement we can do there, we are hoping that will move this in leaps and bounds forward. But, again, we can only measure what we have access to, and it is pretty limited, especially in COPD right now, which is unfortunate.

CO-CHAIR ROSENTHAL: Paul?

DR. BARNETT: I was just going to say we ought to write that one down for when it comes to the final recommendations, that whole idea that the system needs to do a better job of coding and reporting and assessing the costs of the preventive services. I mean the state of coding is pretty abysmal. It is very hard to tell what is going on or what it costs. And to the extent that we can have any impact on the world, that might be --

CO-CHAIR ROSENTHAL: Yes, you also can=t do cost/benefit analysis if you don=t really know what some of the costs are. And there is so much belief about things that are cost-effective, many of which may turn out to be actually cost-effective in reality, but it is very hard to measure.

DR. BARNETT: But there is just like a handful. I am not even sure more than three or four CPT codes to report preventive services.

CO-CHAIR ROSENTHAL: Well, nobody does it, right.

DR. BARNETT: Right. That are psychosocial interventions.

CO-CHAIR ROSENTHAL: Right. Well, we did preempt a little bit because this will be part of tomorrow, but that=s fine.

I would suggest now that we go ahead and call the question on usability, and this is 1, high; 2, moderate; 3, low, and 4, insufficient.

(Whereupon, a vote was taken.)

MS. FANTA: And Doris, your vote on usability?

DR. PETER: Yes. Sorry. High.

MS. FANTA: High. Okay. That=s okay.

It=s 7 high and 7 moderate.

CO-CHAIR ROSENTHAL: All right. Great. So, now we have feasibility, and is there anything left to be said about feasibility?

DR. ELWARD: No, it is very consistent with asthma, and we all voted it very high.

CO-CHAIR ROSENTHAL: Okay. And there are really virtually no differences here. It is coded data. It is what it is.

So, if there is no further discussion, I am going to call the vote on this. And this is 1, high; 2, moderate; 3, low, and 4, insufficient.

(Whereupon, a vote was taken.)

MS. FANTA: And Doris, your vote on feasibility?

DR. PETER: High.

MS. FANTA: Okay. So, we have 10 high and 4 moderate.

CO-CHAIR ROSENTHAL: All right. And now, we are left with recommendation for endorsement or against endorsement. And this is 1 is yes; 2 is no, and 3 is abstain.

Is there any reason to have any further discussion on the overall measure?

(No response.)

All right. Hearing not, let=s vote.

(Whereupon, a vote was taken.)

MS. FANTA: And Doris, your vote on the recommendation?

DR. PETER: Abstain.

MS. FANTA: Okay. So, we have 14 yes.

CO-CHAIR ROSENTHAL: Okay. Who abstained?

MS. FANTA: So, we have 13 yes and 1 abstain.

CO-CHAIR ROSENTHAL: Who abstained?

MS. WILBON: Doris.

CO-CHAIR ROSENTHAL: Oh, okay. Oh, abstain? I thought I heard Athe same@.

(Laughter.)

Sorry. That is what I thought I heard.

All right. We have public

comment. Let=s do that.

MS. WILBON: Hi, Tom. Are you there with us still?

THE OPERATOR: Yes, I=m here.

MS. WILBON: The operator Tom.

THE OPERATOR: Yes, I=m here.

MS. WILBON: Okay. Can we open it up for -- is there anyone on the participant line?

THE OPERATOR: We do have one participant line. Let me go ahead and open that for you.

MS. WILBON: Okay. I guess we could open the line up for that person to make a comment, if they would like.

THE OPERATOR: The line is open.

(No response.)

MS. WILBON: All right. No comments.

So, that will conclude our day today. Thank you all for persevering. It was a little rough; I am not going to lie. But

you guys made it through, and tomorrow, hopefully, will be a little less arduous.

And hopefully, we will be able to kind of bring back some of these ideas and talk them through a little bit and make some recommendations for next steps.

Thank you to the Co-Chairs.

Janet, I am not sure if you are still on the phone, but thank you for dialing in.

Kurt, thank you.

Also, to the Bone/Joint TAP Chairs, who are probably not on the phone, but I just want for the record to thank them for dialing in. And obviously, their input is really helpful.

DR. ELWARD: We had an awesome staff to work with, Tom. Thank you.

MR. AMIN: One other quick thing that I would just like to add. As we sort of think through the structure of tomorrow, I would just want to set a little bit of the

stage.

A lot of the structure for tomorrow is some of the challenges that we noticed, as part of the NQF staff, and then, also, a lot of the challenges that were noticed through the Steering Committee and through the TAPs, evaluating all of the measures, including the ABMS measures, a lot of which didn=t get to this point.

Although there are big sort of methodological questions, some of them are theoretical questions, and they span those two, which is sort of difficult to go back and forth as we sort of go through the module, although we have framed a lot of the big-picture questions along the modules that we have structured the measure evaluation process through.

So, we have posed the questions, and many of the questions don=t really have actual answers right now, considering where the field of resource use measures is. But,

as we are sort of looking forward, after we have gone through this whole process together over evaluating all of these measures, we thought it would be really valuable to sort of harvest a lot of this information of some of the challenges that you have felt in evaluating the measures, some of the tougher theoretical issues that potentially may be out there, noting the limitations of the data that many of these measures use, along with balancing how much we can possibly expect from measure developers who are in this field.

So, with that being said, we just wanted to frame the discussion for tomorrow. And hopefully, we can come in with a good breakfast and be ready for some of these sort of heavier questions, and sort of bear with each other in just sort of expressing some concern or just challenges that we have had to this process of actually evaluating all of these measures.

And we really appreciate any

feedback, if we can get it, from multiple different perspectives, as we sort of think through and advise the community of people who are not only developing these measures, but also the next measure evaluation, as we think through the CMMI potential application and, also, our big lift next year of looking at the public sector episode grouper evaluation.

CO-CHAIR STEINWALD: A question for staff: what kind of feedback, and when would you like the feedback, on the Draft Report?

MS. WILBON: Yes. So, that ship sailed today.

(Laughter.)

We posted the report today for public comment, but you can comment during the comment period. If you still have comments, we can integrate those into the comment process where we gather everyone=s comments, public and members.

We will also be having another

report that will reflect these measures. We will integrate probably a lot of the same ideas. So, if you have any input on how we can improve as we kind of use some of that same information for the second report, that would still be very helpful.

CO-CHAIR STEINWALD: And do you like track changes? Do you like hard copy?

MS. WILBON: You can do track changes, or if you have made hand notes, we will take those, too.

CO-CHAIR STEINWALD: Okay.

MS. WILBON: So, we are not picky.

CO-CHAIR STEINWALD: All right.

MS. WILBON: And we start a half an hour earlier tomorrow than we did this morning, according to my look at the agenda. Correct?

MS. WILBON: And we finish earlier, too.

CO-CHAIR ROSENTHAL: And we finish earlier, too.

So, see everybody in the morning.

MS. WILBON: Thank you.

CO-CHAIR STEINWALD: The meeting is adjourned.

CO-CHAIR ROSENTHAL: Yes.

(Whereupon, at 4:54 p.m., the foregoing matter went off the record.)