NATIONAL QUALITY FORUM

+ + + + +

COST AND RESOURCE USE PHASE 3 PULMONARY MEASURES STEERING COMMITTEE MEETING

+ + + + +

WEDNESDAY
JUNE 25, 2014

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 9:00 a.m., Brent Asplin and Lisa Latts, Co-Chairs, presiding.

PRESENT:

BRENT ASPLIN, MD, MPH, Catholic Health Partners, Chair
LISA LATTS, MD, MSPH, MBA, FACP, LML Health Solutions, Chair
ARIEL BAYEWITZ, MPH, WellPoint, Inc.
LARRY BECKER, Xerox Corporation
MARY ANN CLARK, MHA, Intralign
CHERYL DAMBERG, PhD, RAND Corporation*
NANCY GARRETT, PhD, Hennepin County Medical Center
ANDREA GELZER, MD, MS, FACP, AmeriHealth Mercy Family of Companies
MATTHEW McHUGH, PhD, JD, MPH, RN, CRNP, FAAN, University of Pennsylvania
JAMES NAESSENS, ScD, MPH, Mayo Clinic
EUGENE NELSON, Dsc, MPH, Dartmouth Institute for Health Policy and Clinical Practice
CAROLYN PARE, Minnesota Health Action Group
JOHN RATLIFF, MD, FACS, FAANS, American Association of Neurological Surgeons

JOE STEPHANSKY, PhD, Michigan Hospital Association
LINA WALKER, PhD, AARP - Public Policy Institute
HERBERT WONG, PhD, Agency for Healthcare Research and Quality*

NQF STAFF:
HELEN BURSTIN
ANN HAMMERSMITH
TAROON AMIN
QUINTIN DUKES
ANN PHILLIPS
LINDSEY TIGHE
ASHLIE WILBON

ALSO PRESENT:
BEN HAMLIN, NCQA
LEIN HAN, CMS
NANCY KIM, Yale University
STEVEN SPIVACK, Yale University*
BOB REHM, NCQA

*Via teleconference

TABLE OF CONTENTS

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

9:00 a.m.

MS. TIGHE: Good morning, everyone. Thank you, to those of us in the room, for joining us today for the Phase 3 Cost and Resource Use in-person meeting to discuss the pulmonary condition-specific measures. If you could just begin to take your seats, and I know we have some people on the phone.

We will go ahead and just kind of do a brief run-through of the plan for the day and the plan for tomorrow, some introductions.

Most of you have been here before at this point, but just a reminder that the restrooms are past -- if you exit the conference room, if you go past the elevators, the restrooms will be down the hall to the right.

We will have a couple of breaks today. We'll also have lunch provided at 12:30.

It has been displaying, but if you need the password for the wifi network, the wifi

network is NQFguest. The password is NQFguest. Certainly, we can send that if anyone needs more information.

I'll just go ahead and briefly introduce our project team.

So, we have our new Project Manager Quintin Dukes, who is sitting to my right. He has just joined us about two weeks ago. So, he's coming up to speed on one our more challenging projects, certainly, but glad to have him.

We have Ann Phillips, who is the Project Analyst for the team.

I'm Lindsey Tighe. I'm the Senior Project Manager. I was here in Phase 1 and I'm back for Phase 3. And then, of course, Ashlie Wilbon. Taroon, also will be joining us later. So most faces are familiar at this point.

We'll get the slides caught up to where I am, since I'm reading from them.

I'll just go through, briefly, the agenda for the day. Obviously breakfast has

happened already. We'll have Ann Hammersmith lead the group through some disclosures of interest and introductions.

Ashlie and I will run through a brief project introduction and overview of the evaluation process, and then we'll move right into the evaluation of the candidate measures.

We do plan to get through all three measures during today's discussion. So it will be certainly, a very thoughtful day.

There is an optional happy hour at 6:00 p.m. at Mio, which is very close to your hotel on Vermont Avenue. Feel free to join us. We'll be there. I will just remind you all, as I'm sure you know, NQF cannot pay for this happy hour, but you're welcome to join us.

And then I will turn it over to Ann Hammersmith, our General Counsel, who will do some introductions and disclosures of interest.

MS. HAMMERSMITH: Thanks, Lindsey. I see a lot of familiar faces. So I think you

know the drill, but I'm going to tell you again anyway.

You received a form for us, it was fairly lengthy, asking you about your professional activities.

So, what we'll do this morning is, we'll go around the table, ask you to introduce yourselves and tell us if you have anything to disclose.

We do not want you to summarize your resume. We only want you to disclose things that you believe are relevant to the subject matter that the Committee will consider.

Just because you disclose does not mean you have a conflict. A lot of people think, "If I speak up, it means I'm conflicted." It doesn't. A lot of this is disclosure. It's pure disclosure, so that everyone knows where you are coming from.

I want to remind you that you sit as a individual. You don't represent your employer. You don't represent anyone who may

have nominated you for service on the Committee.

One thing that makes NQF's process somewhat unique is that we are not simply interested in financial disclosures. Because of the nature of the work we do, and that you do, you may have served on committees as a volunteer. No money may have changed hands. Where what you did was relevant to the work of the Committee, we would expect you to disclose that.

In addition, we are particularly interested in grants, research or consulting that you may have done, but only if it is relevant to the work of the Committee.

So, let's go around the table. Tell us who you are, who you're with and if you have anything you wish to disclose. And the Chairs get to start.

CO-CHAIR LATTS: All right, well, then I will start.

Good morning, everyone. I am Lisa

Latts. I am a physician based in Denver, formerly with WellPoint, now independently consulting, and I have no disclosures.

CO-CHAIR ASPLIN: Good morning. Brent Asplin, Chief Medical Officer with Catholic Health Partners in Ohio, and I have no disclosures.

MEMBER GELZER: Good morning. Andrea Gelzer. I'm Chief Medical Officer for AmeriHealth Caritas. We're a health plan. So, one of the measure developers -- all of our health plans are NCQA-accredited. Other than that, I have no disclosures.

MEMBER PARE: Hi. I'm Carolyn Pare, President and CEO of the Minnesota Health Action Group. I sit on the NCQA Standards Committee and do work with Minnesota Community Measurement, but I don't have any conflicts.

MEMBER BAYEWITZ: Ariel Bayewitz. I'm Vice President of Provider Analytics at WellPoint. I used a lot of the NCQA measures for payment models and reporting, but I don't

have anything further to disclose.

MEMBER McHUGH: Matthew McHugh. I'm from the University of Pennsylvania and I'm a health outcomes and policy researcher, and I have no disclosures.

MEMBER STEPHANSKY: Joe Stephansky. I'm with the Michigan Health and Hospital Association, a trade and advocacy organization, and I have nothing to disclose.

MEMBER RATLIFF: Good morning. I'm John Ratliff, practicing neurosurgeon at Stanford, co-sponsored with AANS and AMA. I've got nothing to disclose. And I missed the first meeting because of the snowpocalypse that didn't actually come to be, so I'm glad I made it out here this morning. Thank you, all.

MEMBER WALKER: I'm Lina Walker. I'm with the AARP and I have nothing to disclose.

MEMBER NELSON: Good morning. Gene Nelson from Dartmouth. We have subcontracts around quality measurement with

CMS and Yale and BoozAllen, NCQA is part of that, and working with grants or contracts from PCORI and Robert Wood Johnson Foundation and National Institutes on Aging, all of which go to outcomes and cross-measurement. I don't think there is any conflicts.

MEMBER NAESSENS: Good morning. I'm Jim Naessens, a health services researcher at Mayo Clinic, and I have no disclosures.

MEMBER GARRETT: I'm Nancy Garrett, the Chief Analytics Officer at Hennepin County Medical Center, and I'm serving on the NQF Committee looking at the issue of risk-adjustment and socio-demographic factors.

So, I noticed in the comments, it's going to probably come up today. So, I'm happy to give people an update on where that's at, if that would be helpful at some point today.

MS. CLARK: Hello. I'm Mary Ann Clark from Intralign. Nothing to disclose.

MS. HAMMERSMITH: Okay, thank you.

I understand there are some Committee members on the phone. I will call your name. Cheryl Damberg.

MEMBER DAMBERG: Yes, Cheryl Damberg from RAND. I have nothing to disclose.

MS. HAMMERSMITH: Okay, thank you. Jennifer Eames-Huff? Is Jennifer Eames on the phone?

(No response.)

MS. HAMMERSMITH: Okay, Martin Marciniak? Is Martin Marciniak on the phone?

(No response.)

MS. HAMMERSMITH: Herbert Wong?

MEMBER WONG: Yes, Herbert Wong with the Agency for Healthcare Research and Quality, and I have nothing to disclose.

MS. HAMMERSMITH: Okay, thank you. I just want to remind you of a few additional things.

We sent you, in addition to the general disclosure of interest form, you were given measure-specific disclosure of interest

forms. My understanding is that no one has any conflict there. If you think you do, or something has changed, please speak up when that measure is considered.

The last thing that I want to remind you of is that we rely on all of you to help us make our conflict of interest process effective.

So, if you were sitting there, you think that you may have a conflict, you think that one of your fellow Committee members may have a conflict, you think somebody is acting in a biased manner, please speak up. We don't want you just sitting there thinking, "This isn't quite right," and not telling us.

So, if you do want to raise anything like this, you can always raise it openly in a meeting. You can go to your Co-Chairs, who will go to NQF staff. Or you can go directly to NQF staff.

So, based on that, based on the disclosures this morning, do any of you have

anything that you want to discuss regarding disclosures or any questions of each other?

(No response.)

MS. HAMMERSMITH: Okay, thank you.

MS. TIGHE: All right, thank you, Ann. So, moving onto the next slide. Okay, we just wanted to give a brief overview as to where we've been and where we are currently.

Okay, so, this project is a three-phase effort, as many of you are well aware.

In the first phase, we looked at total cost, non-condition specific per capita or per hospitalization episodes. One measure was endorsed during this effort. It was the Medicare spending per beneficiary measure.

In Phase 2, we looked at cardiovascular condition-specific per capita and condition-specific episodes. As you have been updated over email, those three measures were recommended and they're out for NQF member vote at this point in time.

And then today, we're focusing on Phase 3 of the pulmonary and other condition-specific per capita and condition-specific episodes.

The next slide, we've got just a little bit more detailing of the three -- oh, is there one before this?

MS. WILBON: Just give us a second. I think this may be a different version of slides that we uploaded. Just give us a second to switch this over.

MS. TIGHE: Okay, well, I'll just speak to them while we're looking at the slides.

So, as I mentioned, all three measures that were submitted to Phase 2 were recommended for endorsement.

The first measure, 1558, which was the relative resource use for people with cardiovascular conditions, was the NCQA measure that was recommended for endorsement during the in-person meeting.

2431 and 2436, which were the

Yale/CMS measures for hospital level risk-standardized payment for AMI and for heart failure, were initially in the gray zone, where consensus was not reached during the in-person meeting. After reconsideration on the post-comment call and some additional discussion with the developer, these measures were re-voted on after the commenting period.

The AMI measure was recommended with 14 'yes' votes and seven 'no' votes, which would be a 66.67 percent approval rate, which puts us into the recommended for endorsement.

2436, the heart failure measure, it was recommended by 13 Committee members and not recommended by eight. At 61.9 percent, it squeaked by into a recommended measure.

So, all three of these have gone out for vote as recommended for endorsement by the Committee.

As you know, they're out for NQF member voting. So, we'll see where that comes back, if we reach consensus with our membership

or not.

If we do reach consensus, it will go on to CSAC consideration during their July 9th meeting. If we don't reach consensus, we'll be going out to our stakeholder councils, getting some more input, understanding the issues and having some rehashing, similar to what we did for the Medicare spending per beneficiary measure.

So, that's kind of the update on Phase 2. Are there any questions or comments before I move on?

(No response.)

MS. TIGHE: Okay, so, moving into Phase 3, we do have three measure submissions today.

There is the relative resource use for people with asthma, relative resource use for people with COPD, and then the hospital level risk-standardized payment associated with a 30-day episode of care for pneumonia.

The first two, the relative

resource use measures, are maintenance measures. They were recommended for endorsement during what was actually the first phase of this work, which may not have involved most of this Committee, but was several years ago.

Just walking through the timeline. At this point, we're at our in-person meeting, of course. This draft report, as we come out of this meeting, will be posted for NQF member and public comment August 14th through September 12th.

You all recently should have received a calendar update with a new appointment for the call to review and respond to comments, which will be on September 24th.

And then the draft report will be posted for NQF member vote in October, will go our Consensus Standards Approval Committee in November, and then endorsement by the NQF Board of Directors during late November or early December.

Okay, I will turn it over to Ashlie for an overview of the evaluation process, unless there are any questions.

MS. WILBON: I'm just going to pause for a second give Ann a chance to upload the slides. I think it will be easier for you guys to follow along if you have something visual to look at.

But a lot of it is the same material. It's really just a refresher. I tried not to regurgitate the criteria, but really just give you some key guiding questions for things you should be thinking about as we evaluate each of the criteria.

We have given you the algorithms that we had you guys working with on the last meeting, which caused a lot heartache, which we recognize.

So, it is in your packets for reference, but, you know, if you have kind of questions or you want to kind of see what types of things you should be thinking about as we're

having the discussion around the criteria for each of the measures, it's there for your reference. But we're not going to structure the discussion so much around the algorithm, per se, but we will kind of refer to it here and there to make sure we're staying on track with what we should be evaluating for the criteria.

So, we heard you guys and we recognize that operationalizing isn't always as easy when you're in the moment. So, we'll do our best to guide you through that process.

And, as always, you guys have been doing this for a while, but if there are questions and clarifications, you can certainly do that along the way.

The other thing that we wanted to do to kind of start this phase off is, because you guys have been reviewing these measures, you're getting so good at it at this point, the measures that we reviewed in Phase 2 are from the same developer. They essentially used the same methodology to structure the measures.

And we've already had a work group call, as well, where you guys had an opportunity to discuss some of the information that the Technical Expert Panel came up with, and have a little bit of early discussion about the measures. Which was also something that we integrated as a result of the feedback that you gave us last time in wanting to get a little bit more into the measure before the meeting, so that you could kind of have a little bit more thoughtful discussion at the meeting.

So, hopefully you found that helpful. And so we'll pull up a slide shortly that will show some of the main issues that we've encountered over the evaluation of these measures, and just have a little bit of discussion on where the Committee stands on some of these issues, so that we're not kind of rehashing the same issues over and over again.

There may be some kind of broad agreement that the Committee has on some of these issues, like hospital transfers, who

accepts responsibility for cost for patients that are transferred from Hospital A to Hospital B. Where we stand on some of the attribution and some of the risk-adjustment challenges.

And some of these don't have answers. We recognize that, and we've gone back through the reports to kind of see how we've characterized the Committee's responses to some of these other issues that have come up before, and there aren't really answers.

But we just kind of want to have an open discussion about some of these issues, as we already know that, given the similarity of these measures to the measures we've already discussed, that a lot of these issues will come up again. And so maybe there is kind of some broad agreement that we have on how the Committee would like to handle these issues as they come up in the measures, so that our discussion, when we get to the actual measure, is a little bit easier.

So, with that, I'll just point out some of these issues. Again, I pointed out the approach to transfer patients. We've talked about attribution, risk-adjustment.

In the last phase, there was concern again, particularly with the Yale/CMS measures, around the R-squared value, whether or not it accounted for enough of the variation.

We've also had discussions around SES, which Nancy has graciously offered to provide the Committee with an update on where that Committee is. But we've kind of put that on the back burner for now, until we have some additional guidance.

We are kind of working with the criteria that we have for now, and NQF is not changing the risk-adjustment criteria at this point, until we've seen this other kind of parallel effort through.

The other issue that has come up is the handling of deaths and whether or not the cost for death should be included or excluded.

Pharmacy data has been another issue. Some of the measures have not had the capability, but I think particularly with the CMS measures, of including Part D claims, which are for pharmacy.

And the linking quality and cost issue has always been an overarching issue with these measures, and we do have also another parallel effort going on with that, as well. We'll spend a lot more time talking about that in the agenda that's planned for day two, and to give us some guidance on where we go with that piece.

But just with that brief overview, I'd like to kind of just open it up and hand it over maybe to Lisa and Brent to gauge whether or not the Committee has any thoughts on these issues and how we might broadly approach the discussion on some of these issues as we move forward.

CO-CHAIR ASPLIN: Okay.

MS. WILBON: Sorry, or if there are

additional things that you think should be on this diagram that we haven't included, feel free to add those, as well.

CO-CHAIR ASPLIN: Thank you, Ashlie. So, we have the slide in front of us. Are there any of the bubbles that you see here, the different issues that Ashlie just walked through, that any of you would like to have a conversation about right now?

CO-CHAIR LATTS: Or is there anything that's missing that we've had issues with?

CO-CHAIR ASPLIN: Some of these, to me, end up falling into a couple of different general categories.

Like for example, the transfer issue seems to be sort of almost a policy issue, as a Committee, that we settle on one way or the other and deal with.

Pharmacy data, on the other hand, may be something that we ask for in future measures, given that, currently, the

variability in enrollment in Part D and the reporting of it would unfairly penalize some hospitals and not others.

So, it's really not fair to put it in now. But perhaps, over time, if there is any other way to manage that, it may be something that would be of interest.

CO-CHAIR LATTS: In looking at this, it sort of strikes me that there is a bunch that are around the inputs. There are several that are around what=s done and how the data is treated, and then there=s the outputs. You know, what do you with the information and how do you use that? And we've had a lot of questions and concerns about that in these meetings, although it's sort of come up repeatedly that there is not much we can do about it, essentially, besides be as true as we can to the measure and then put it out there and see how the universe uses it.

MEMBER GARRETT: So, I'd be interested in a quick summary of the linking

quality and cost work. I haven't had a chance to read the draft report. Would anyone be able to give us a summary?

MS. WILBON: Sure. So, that's an effort that we've been working on, again, parallel to this work that's funded by the Robert Wood Johnson Foundation.

We convened some experts. I think some of you guys were on that Committee. Maybe not. There are some people who are on this Committee who are serving on that piece.

Wait, Joe, you're on that, right? The linking cost and quality.

So, feel free to -- and Herb, as well. Thank you. So, feel free to chime in if I miss anything, or you'd like to add.

We basically convened the experts to really think about some of the methodological challenges around actually combining cost and quality measures, and also thinking about some of the pros and cons or the different approaches that are out there.

We asked the authors, Andy Ryan, who is actually on this Committee, and Chris Tompkins were the authors, and they did a really nice environmental scan of some of the approaches that are out there in the field for different entities who are combining cost and quality signals in different ways, and characterizing them so that there is some order or some understanding about the pros and cons of the different approaches, how efficiency is measured, what the thresholds are set at. And, really, in the different ways that the measures are linked, it gives you different types of information.

You know, if you set the quality threshold at a certain point and then compare on cost, or if you set the cost threshold and then compare on quality, whether or not the providers are efficient, they really end up in different quadrants. So, kind of comparing the pros and cons and laying out some of the differences in the approaches.

They also provided some operational guidance for NQF on terms of where they would recommend that we go in the future, in terms of evaluating the linkage of cost and quality measures. And that's really something that we're going to spend some time talking about tomorrow, because, really, we kind of feel like that is the foundation of what this group was initially established to do.

We, initially, for those of you that have been with us from the beginning, this was called the Efficiency Committee. And that was because, ultimately, the goal for us always was to get towards efficiency and value, but we really hadn't gotten to the point yet where we were ready to say what we were asking for when we were saying we were evaluating efficiency measures.

So, this report has really helped us, I think, get a lot closer to doing that. But I think there is still probably a few questions we'd like some input from you guys on, and we

need to think through a little bit more about how we operationalize that, what is it that NQF will be endorsing, do we change our current process for evaluating resource use measures by asking for some additional information about how those measures would be linked to quality.

You know, so, there is a lot of there. We're still trying to work it through and see what the implications are for our process. But if you get a chance tonight, I would really suggest reading the paper. It's a good read, and I think it will help the discussion tomorrow, as well. So, does that help, Nancy?

MEMBER GARRETT: That's very helpful. Would you send us the link, so that we could --

MS. WILBON: Yeah, it's in the discussion guide. If you go on SharePoint, I think it's linked in the discussion guide, but we can also send out the actual report, as well, if it's easier for people to just get to.

So, is everyone settled kind of on that bubble slide? There didn't seem to be a lot of --

MEMBER GARRETT: So, if we don't talk about those issues now, we can't bring them up during the day?

MS. WILBON: No.

(Laughter.)

CO-CHAIR LATTS: That=s right, this is an all-inclusive list of everything we can discuss.

MS. WILBON: Not at all. We were trying to create some efficiencies, if there were some to be had. But you guys are still asleep a little bit, so that's fine. I know you guys will get warmed up as we go.

So, the last point I just want to raise is that, you know, we want to make sure, as much as possible, that we are remaining internally consistent. I think because these measures are very similar to what we've done before. There is the possibility, you know, as

discussion evolves, that decisions are made that may be contradictory or vary from where the Committee has landed before. And that was kind of also our effort at kind of laying some of the issues out, to make sure that the Committee had an understanding of where they land on some of these issues and that they can be consistently applied in the evaluation of the measures.

So, what we'll do is, throughout the day, we'll just make sure that, you know, if there are any check points, and staff and the Chairs will be working together to make sure that we're staying internally consistent and kind of, you know, raising any questions, if we feel like things are veering in that direction. So, does that sounds fair? Okay.

I'm sorry, Lina?

MEMBER WALKER: Hi, Ashlie. Just a quick follow-up comment to the point you made about consistency.

I understand that the three measures, the three Yale measures, were kind of

created and developed very similarly, using the same methodology.

But the conditions are different. So, sometimes, even though the methods are the same, the outcome can be different because, you know, one condition might have a lot of variation in severity, for instance. So, the lack of an adjustment for severity might be more of an issue for one measure, rather than another.

And then, you know, things like not having pharmacy cost. The pharmacy cost is a really large input in the total cost of care for particular conditions. You can imagine then that might be more of a deficiency for one condition rather than another.

So, I just wanted to mention that, because there are reasons why, you know, votes might be different even if the methodology were the same.

MS. WILBON: That's a very good point. Thanks, Lina. And we certainly

recognize that. I think as long as there is, you know, adequate discussion and there is justification for that, I think we're perfectly fine with that. Anything else?

Okay, next slide. So, here are four criteria. We'll focus, again, on importance to measure and report, scientific acceptability, feasibility, and use and usability.

Again, we're not going to spend a lot of time on these, but, again, the goal is to make sure that the topic of the measure is important to measure, that there is potential for driving improvements.

Scientific acceptability is really focused on ensuring that the measure is reliable and valid.

Feasibility, we want to make sure that there is not an undue burden to implement the measure.

And with the usability and use criteria, we're looking to make sure that the

outcome or the results are usable. And in particular, with cost and resource use measures, we're also looking at the issue around transparency, the ability to kind of understand what's behind the measure.

MEMBER BAYEWITZ: In terms of is the measure important, you know, a lot of the RRU measures focus on rolling up at the plan level.

So, if we feel as though that isn't a good way to evaluate, would that go at the use? Or would that go as that -- the first -- could you just go back a slide? No, the one before that. The one that was showing -- yeah, there you go.

So, important to measure import is a must-pass, whereas usable is not a must-pass. So, if the way that they're using the measures to evaluate plans, is that going at usable or is that going important to measure?

So, theoretically, that measure could be used -- you could roll it up at multiple

levels and one could be important to measure for a certain purpose, but the way that it's being submitted, you know, we might feel that's not the case.

MS. WILBON: Yeah, I would say the importance criteria is more focused on the topic areas.

So, for example, with the RRU measure, they're measuring total cost for asthma patient for the course of the year.

So, the question would be is it important to measure all the cost that an asthma patient would use for a health plan over the course of a year, and is there room -- or is there potential for driving improvement in that area?

So, it's more about the topic area, and I would say less around the actual approach, which may actually go more into the validity and the usability criteria, in terms of how they've actually structured the measure to measure that, and whether or not that is the

right approach. Does that make sense or no?

MEMBER BAYEWITZ: It does. So, if you feel as though it's important to evaluate asthma costs, right, to look at that in terms of giving feedback to a provider, let's say a large organization, do you feel as though that is helpful beyond quality, right?

But you feel as though when you roll it up to a plan level, it's not -- it's not important because it's not actionable. There is nothing that the plan can do that's clear.

Is that -- is it no longer important, because that's the way that they're currently submitting the measure or is it still important because one could theoretically look at that measure and use it for something else?

MS. WILBON: I would say -- I don't want to split hairs here -- but I would say the latter part, in terms of whether or not it's actionable really is kind of a usability discussion.

MEMBER BAYEWITZ: Okay.

MS. WILBON: And then we're trying to kind of parse out with the importance of whether or not, you know, it would be important to measure costs for asthmatics over the course of the year and whether or not there is enough variation in cost for those patients that, measuring cost in that area would, in some -- somehow illuminate variation or help us to improve the variation and cost in that -- for those patients. Does that help?

MEMBER BAYEWITZ: Okay.

(Off microphone comment)

MS. WILBON: Okay, does that help?

MEMBER BAYEWITZ: Yes, it helps.

MS. WILBON: Carolyn you --- okay, you have this? Okay, I didn't know if you were giving a look like, "That didn't make sense."

MEMBER BAYEWITZ: No.

MS. WILBON: Okay.

MEMBER PARE: I'll comment on the look, since I'm making the face, and I just think it's an interesting thing to think about

because as I've looked at these cost and resource use measures and the issue of importance, there is a number of different ways to look at it, because so much depends on for what purpose and by whom, and that's not globally applicable.

MEMBER BAYEWITZ: Right, and just as a bubble, to add to your bubble, I mean, I think just from speaking from a plan perspective, the more cost measures that we can have that are in fact actionable, that would be helpful, and I think even though we're not saying it's a must-pass, I mean, if we're only looking at three measures every, you know, six months or whatever number of months, and we're saying -- and maybe as a group, we don't all agree about this.

But if we do say, "This is not actionable. There is nothing you can do with it," it is technically, you know, scientifically acceptable, and why -- just from a prioritization standpoint, why would we put

those four, I mean, just into in terms of spending out time?

CO-CHAIR LATTS: I think that also starts to get to the issue which we've touched at in previous meetings and I suspect we'll get to tomorrow, which is that we're not really getting the measures we want.

You know, we're reviewing and spending a lot of time reviewing the measures that come our way, and you know, we've had this in basically, every single Panel I've been on, which is that NQF is essentially a passive process.

We've put out a call for measures and then we review what we get in, without the ability to say, "These are the measure we want," and nobody is developing them, and so, I think it's an ongoing frustration of how do we get more measures that are aligned with what -- you know, at least the Committee, and there is a fair amount of expertise on this Committee, the measures that we want?

CO-CHAIR ASPLIN: Larry?

MEMBER BECKER: so, Lisa, I agree with you. The question is do we have a list of the measures we want, and can we create that and put that out, so that when people are thinking about it, they develop the measures that we want?

CO-CHAIR ASPLIN: Gene?

MEMBER NELSON: I agree with both of those comments that were just made, and perhaps tomorrow is a chance to specify the measures that we want.

Getting back to the health, we're warming up, I think.

CO-CHAIR ASPLIN: We'll get to the measures by noon.

MEMBER NELSON: Getting back to the health plan level and is that actionable or not.

I find that Dartmouth-Hitchcock health system and we have a lot of asthma patients, which we do, and if the health plan provides us with information that tracks total

cost per year, that sometimes is in our system and sometimes is out of the system, that's helpful to us.

I would say that could be actionable. We can bring it down to the level of the health system. We can look at the care teams that are responsible for those asthma patients. We can look at their patterns of care, and if they're out of line, higher cost and lower benefits, we can do something about it.

MEMBER BAYEWITZ: And like I said, I could see why an RRU would be helpful to a provider system, and from a health plan perspective, if I could effectively quantify RRU's at a system level and whether that statistically is meaningful, I don't know.

I could understand that, because providers could actually drive a lot of these things and providers could evaluate whether certain resource use makes sense or not.

Again, we don't know if more is

necessarily bad in certain cases. But the way that it's being used, at least to my understanding here, is rolling it up at a plan level. It's not looking at Dartmouth, but looking at an overall, you know, Aetna versus United, and saying at a plan level, or WellPoint or some other Blue, and saying at a plan level, what is the observed versus expected? You know, who is "better" or who has higher use? And to me, that -- I don't know if it's actionable. I don't know what you do with it.

I mean, some of these, I mean, we'll get to it when we talk about, you know, COPD, but you know, 40 percent of plans moving two quartiles, you know, for an observed versus expected measure, I mean, what do you do with that?

So, again, that's why I'm wondering -- I could see it might be important, if scientifically we said, "It can be evaluated at provider level," and then I would say 100 percent, but if I'm not seeing that, if what I'm

seeing is more of the scientifics around at a plan level, to me, that, I just don't -- I don't see why an employer, for example, would, you know, care as much.

You know, to me, they would care about what is the quality and they would say what is the actual cost, but a resource use, which is where you strip out unit costs, I don't know what they would do with that.

CO-CHAIR ASPLIN: Helen?

DR. BURSTIN: Good morning, everybody. Helen Burstin. Sorry I was late. Larry and I were having a chat this morning about consensus with Chris Cassel.

Just a quick reflection on the gaps piece, and this does come up pretty often.

I think one of the challenges -- there are several challenges here. I think one of the biggest challenges is, we often times have lists of measure gaps that are frankly, not specific enough.

I mean, people just put out lists

saying, "We need more cost research use measures." That isn't enough really, I think enough of a signal for the developers to really kind of dive in.

The second thing is, we do still have significant data limitations.

Now, a lot of what you see being brought forward to you is based on what is available at this time, in terms of data. So, I think some of this is also thinking prospectively about how you could use different data to bring it together from different sources, to make some sense of this.

So, I think for example, the pharmacy conversation is an important one, of how you begin pulling in those streams of pharmacy claims data, and the third piece of this is that we also need to start thinking about what is actually being used in the field, perhaps more local or regional initiatives or state initiatives, that have been useful, that have moved the needle and start prospecting for

those measures.

I was just on a Panel with Arnie Milstein, Dana Safran and Jill Yegian from IHA at AcademyHealth, and it was really interesting to see the cost measures that IHA and each of them use.

So, I think we need to increasingly start looking at what's being used on the ground as a starting point, rather than assuming everything has to be built de novo from the ground up in terms of new measures.

So, this is where we really look to you, as you know where there might be good examples of cost resource use or even optimally efficiency measures that could be brought forward and that hopefully, work with our measure developers to see if that looks promising, would they want to try to move that into being a national standard.

But again, it's very easy to be negative about the measures before you, but keep in mind, they're there mainly because of

data limitations. This is the best they could do to date. You know, none of the developers before us have been static. They've all tried to improve on these measures and make them better and better.

So, just keep that in mind, that these are often times based on best available at this time, and if you want to push further and say what we need, be very specific, would be my recommendation.

MEMBER BECKER: So, all of these things are really important, and they are about signals, I believe, to patients and providers about where to go, but they'll never be perfect.

I don't know if anyone has ever run a sales organization, but you put out an incentive plan and the sales people figure out how to game it. I mean, it's like instantaneous, right? They figure out how to work the incentive plan.

So, I think the best for us is to create measures and signals that are

directionally correct to move people in the right direction, at least for my health, that's what I want.

I want a signal. I want a measure. We're never going to get it perfect, but we've got to move to something that is better than a random walk for patients and providers, so they can prove the system, so we can get better care.

So, we focus a lot on the details, but directionally, I think it's really important to move some of these things forward and understand their implications.

MEMBER GELZER: Thanks, Brent. This is a journey. For some of us, this is a long, painful journey, but that said, and one of my comments about the relative resource measures was the issue of unit cost.

But I think from a scientific -- I mean, if the measure is valid from a measurement perspective, and we get to price transparency of some sort and get through all that proprietary stuff, if you add the actual cost

of the drugs and you add the actual costs at each hospital or for each DME or for each service into the measure, they become very powerful and actionable.

So, I agree with you. This is -- you know, this is a step in our journey, and I don't -- I think we need to make it very clear that we expect the journey will accelerate and continue, and we will get additional information, but I think these measures are certainly good enough to go forward.

CO-CHAIR ASPLIN: John and then Gene.

MEMBER RATLIFF: It's something I'm having an issue with, with evaluating these, especially the asthma measure. How do we reconcile changes in NQF policy, while these measures have been developed?

Like now, we have risk-adjustments for socio-economics, which hopefully we'll be talking about. That, to me, seems like it's only high-impact for that asthma measure, and

yet, the developers who are working on this measure were doing it kind of in parallel with the NQF process of developing the risk-adjustment strategy, which we have now.

Do I ding the developer because they didn't bring in socio-economic factors with regards to evaluation of asthma, which I think would be pretty high impact, or do we give them kind of the benefit of the doubt because they weren't party to or knowledgeable of what NQF has now endorsed, or NQF is developing, I should say, with regards to --

CO-CHAIR LATTS: And in fact, historically, they've been specifically told not to adjust for SES. So, they weren't allowed to.

DR. BURSTIN: One comment on that. The report is still -- has not yet been approved. So, we're still acting on the current state, and it's still, as Nancy knows well, still being edited up to the very last moment.

CO-CHAIR ASPLIN: Gene?

MEMBER NELSON: On the issue of level of analysis and aggregation, one of the points underneath the statement that I made, it might be consistent with what Larry Becker was saying, is this idea of parsimonious, powerful set of measures that go to value, and that create the conditions at lower levels, that provide the opportunity for innovation and improvement, without overly -- being overly restrictive.

So, if we have a parsimonious set of measures at the plan level or the health system level, and allow more detailed measures that get drilled down and used internally and operationally, we might be in a better position, because we can easily have -- go from 1,000-plus measures to 10,000-plus measures with different units of analysis being prescribed, rather than creating the conditions for the organizations to work inside, to deliver better outcomes at lower

real costs to society.

CO-CHAIR LATTS: I am just sort of reflecting on this conversation and Helen, what you had said about finding out what's on the ground and in being specific, in terms of what we at the Committee want, and it seems like there is a real disconnect between sort of what we at the Committee want, what's going on, on the ground and what's going out to developers.

I mean, just that process is not -- we're not getting -- we on the Committee are not getting the specific information of what the plans and the providers are using at the ground level. That's not going out to developers and then frankly, there is not sort of the resources to fund both at the NQF level and at the developer level, because the plans and those using these, I won't call them ad hoc, but not -- non-NQF approved measures don't have the resources that the developers have.

I mean, it's incredibly expensive, as we know, to get a measure developed in the

way that -- through the NQF process.

So, there is just a real disconnect in this process and I don't know how to change that, but it's not working.

DR. BURSTIN: We agree completely. We really do need to think about how we have better feedback loops from what's happening on the ground, what's being used, what's moving the needle, what's not. It's been a frustration in how you find it and how you fund it, and frankly, from the developers point of view, there is not a whole lot of money for measure development, as our friends at the table at NCQA over there can attest to.

It is still really difficult, and you're often times being asked to develop a measure for a specific intended purpose, that you believe is funded by CMS, which again, limits the applicability to a specific intended use.

So, it's a challenge, but you know, I think we're hoping through a whole series of

initiatives, to see if we can make more sense of that, particularly in this phase, where I think everybody really wants more, and try to get a handle on that.

CO-CHAIR ASPLIN: The silver lining in that phenomenon is to Gene's point, that it has somewhat slowed down what otherwise, would have been an exponential explosion of measures.

Now, there is still an explosion of measures, right? But the barriers to entry to come through this process have been raised significantly and it has both positive and negative effects, I think, and so, we'll see how that unfolds.

You know, I think if we're going to put our two days together into a frame work, I think this conversation, prior to us getting into the measures is helpful, and then spending the rest of this day to kind of work through some of the similar issues that we've encountered previously with the similar constructed

measures, I think it's going to really help set up a rich conversation for tomorrow, about what we will look for over the long run, and I am still interested in trying to understand how we measure population-based acuity-adjusted healthy base and the cost to produce health, rather than the cost to produce care, over time how can that paradigm shift? Not that we would abandon measuring the cost of the care we provide, but if our ultimate goal is to improve on an acuity-adjusted basis, the health status of a population would be very interesting to understand.

What partnerships of payer and delivery systems can provide the most acuity-adjusted healthy days at the lowest cost for populations? Larry?

MEMBER BECKER: Just a thought, and maybe it's a slight turn off the topic, but in the research world, there is this debate around randomized control trials and observations studies and that, and I wonder if for us, what

we're doing is the RCT paradigm and out in the world, there is the observational piece and maybe we should figure out a way to do both, and spend some time trying to figure out how do we take what is actually in use, with their actual results, and bring that forward, because as opposed to trying the other way.

Right now, we're creating measures and putting them in the field. Why don't we take what's in the field and figure out if that can be spread?

CO-CHAIR ASPLIN: Let's take a moment here to see if either anyone else has joined us on the phone or if either Cheryl or Herb have comments.

Jennifer, have you joined us on the phone?

MEMBER DAMBERG: I don't think Jennifer has. This is Cheryl.

I think that this conversation is really a useful one to have and I think it both helps us get clear on what our expectations are,

as well as sending signals to the field.

So, I'm looking forward to tomorrow's conversation.

CO-CHAIR ASPLIN: Great. Herb, any comments?

MEMBER WONG: No, I think that I agree with Cheryl, that this conversation has been incredibly helpful. So, I have really nothing else to add to the conversation at this point.

CO-CHAIR ASPLIN: And Martin, have you joined us on the phone?

So, not to dive way into the weeds, but we've got to dive way into the weeds here. Cheryl and Herb, have we figured out a mechanism for when you have a comment or a question, as we move through the day?

Okay, you can raise -- we will be looking at the chat. If you want to raise your hand on the chat, and we'll get you into the queue.

MEMBER DAMBERG: Okay.

MEMBER WONG: Very good.

CO-CHAIR ASPLIN: And Ashlie and Lindsey, can you remind all of us where we are vis-a-vis the quorum and how are we going to actually vote, or what are we going to do?

MS. WILBON: We're going to get to that in just a second.

Let me just -- we've actually talked a little bit about our -- or a lot about some of these things. So, I'm going to keep going, and Ann, if you could go to the next slide for me.

This again, it's kind of the question Ariel, that I was trying to get to, but for importance, we're really asking will this measure make significant contributions towards understanding healthcare costs in the clinical area. So, again, kind of at the topic level, and whether or not the developers demonstrate that there is variation and an opportunity for improvement in that area, by using this measure. Next slide.

So, for scientific acceptability, just again, highlighted some key questions or things for you guys to consider as we're discussing these elements of the criteria.

Particularly for reliability, whether or not the measure specifications are ambiguous, small same size, rare event, other random areas like missing data, whether or not those are occurring.

Threats to validity that should be considered or conceptual or clinical mis-alignment with, you know, expected clinical course or something like that, or evidence. If the measure is unreliable, it can't be valid.

Whether or not the exclusions are appropriate. Whether or not there is differences in patient mix, which may attribute a little bit to the risk adjustment and the adequacy of that.

If there are systematic or missing incorrect data, and then also, consideration of

the costing approach and whether not based on how they're attributing dollars makes sense in the -- based on the measure intent. Can you go back one slide?

So, in terms of testing, just want to highlight again, we're asking you guys to consider whether or not the appropriate method was used, whether or not the scope of testing was adequate.

In particular, a reminder that face-validity is our minimum requirement for the measure to pass. We're just asking that the developer has demonstrated that there -- face-validity has been systematically assessed and that through that, that they've demonstrated that it was valid, based on their assessment. And if you guys determine that the measure for that particular criteria, in terms of rating, would be at least a moderate rating for that.

So, I know we've had discussions in the past on whether or not their testing was

adequate and whether or not the face-validity demonstration has been adequately met, and so, just want to remind us of where we are there.

Whether or not the results were in acceptable norms. Again, this is one that's come up around the R-squared value, with the risk-adjustment of the risk model. Again, whether or not the risk model has been adequately calibrated.

So, you guys are obviously on the right track, but considerations that we should -- in terms of conversations we should be having, as we go forward.

In terms of feasibility and usability and use, again, the key thing for feasibility, particularly because these measures don't have any costs associated with them, we're just asking whether or not there is undue burden -- undue burden would be imposed, in order to implement the measure.

That tends to be a pretty easy criteria to get through, so we probably won't

be spending a lot of time there.

Usability and use, I think is somewhere where we will probably end up spending a lot of time, and in our previous discussions, I think -- and a lot of our other committees, and Helen may have comments on this, as well, is that, you know, and there is discussions at other levels of NQF and whether or not endorsement should be considered for certain -- for the measure to be used for certain purposes or certain applications.

We're not there yet. But we do, in this criteria, ask you to consider how the measure will be used, if there is a plan for how the measure is going to be used or in the application that is currently used, is it used in an accountability application, particularly for payment or public reporting, and then the other question is about whether or not the benefits outweigh the risks for unintended consequences. I think that is a conversation that we've had before, as well.

Demonstrating whether or not the performance has improved through the use of this measure. So, through implementing the measure, is there evidence or data that the developer has been able to submit, particularly for maintenance measures, as this is something that we tend to rely more on, to see whether or not there has actually been improvement in care.

So, is there a usefulness in an application implementation of this measure, such that we're getting better information, performance is improving over time?

Then the last piece that I mentioned earlier is around transparency. Particularly with these measures there is a lot of very dense information around, you know, risk adjustment and the specifications and the costing, and due to our multi-stakeholder audience, we want to make sure that there is a level of transparency, particularly for those who are being measured, and those who are using the measure, that that

information, that result, can be kind of broken down for other stakeholders to be able to understand. So, that's another consideration for you guys to think about.

And I would just add to the -- to Ariel's point, which is not on the -- particularly on this slide, but around whether or not the measure is actionable and usable for those who are going to be using the measure or being measured by the measure.

So, in terms of today's meeting, a lot of these things are kind of a given, but being prepared. Hopefully, everyone has had an opportunity to look at the measures and have something to contribute to.

If you have to take a call, an urgent call, we understand, but we do have -- we have strategically placed breaks. We will be very -- we'll be making sure in taking care to make sure that we do take those breaks on time.

We realize that you guys have lives and jobs, and we want to give you an opportunity

to handle any business, but as much as possible, particularly with our quorum issues today, we'd like as much as possible, to have everyone in the room for as much of the meeting, particularly during the measure voting, so that we have all the votes that we need to be able to move the process forward.

We've talked about kind of keeping comments concise and focused, and where possible, being efficient about our discussion, so we're not repeating information, being courteous of each other, allowing others to contribute and I think that's about it.

So, I think Lindsey, is that it?

MS. TIGHE: Okay, so, just as a reminder, for the process forward, discussing the measures, we'll ask our developer colleagues to give a brief introduction to the measure.

We'll have NQF staff provide an instruction to the criterion that we'll be

discussing, and then we'll ask the lead-discussants to provide an overview of the measure and a description of any submission information, comments that were received during the pre-meeting comment period or from Committee members, and have them just kind of provide an high-level overview of where the preliminary ratings and comments came in, and then open it up for Committee discussion, which Lisa and Brent will facilitate.

We'll ask you after each discussion of criteria to stop and vote. A little bit more on the voting process in a few minutes, but from there, we'll ask for the co-chairs just to provide a brief summary of where the Committee landed in their discussion, so that we can ensure that it aligns potentially with the votes, and that we have enough to substantiate our draft report, as we write this after the meeting.

The next slide, just a reminder for the lead-discussants, hopefully this is not the

first time you've seen this, if you a lead-discussant.

We'll have Nancy and Joe for 1560, Matthew and Mary Anne for 1561 and Janice and Jim for 2579.

Andrea has joined our TEP in their discussions, and so, if there are any questions that relate to the TEP evaluation, which has been provided to you in the worksheet or anything related to the clinical specifications for the measure, if it was discussed, hopefully Andrea is able to answer it. If it wasn't discussed, certainly she can let you know. So, next slide.

Voting guidance and process. So, next slide.

As you all know at this point in time I'm sure, measure is recommended for endorsement by the Committee when the vote on the must-pass criteria and the overall vote is greater than 60 percent. It's not recommended when it's less than 40 percent, and then we have

the -- kind of the gray zone, where we haven't reached consensus, if the vote falls between 40 and 60 percent.

If we're in -- this will be a little bit different. We will discuss each criteria for the measure. Going into the next slide.

Okay, just to finish. Sorry, I've got these in the wrong order for myself, but just to finish on this point.

If consensus isn't reached on the measure, we'll put it out for NQF member and public comment. After the comment period, we'd then ask you to reconsider the measure and all of the comments that we received at that point in time, similar to what we just did for the Phase 2 measures.

Next slide. Here is what I wanted to talk about.

So, as you may have noticed, we're a little bit light on Committee members at this meeting today. We have 14 of you in-person. We have two or three who joining by the phone.

NQF has defined a quorum as 75 percent of the Committee members, which would be 18 members for us.

So, we're not quite at a quorum of Committee members participating. So, to address this and to ensure that we have a thorough discussion of the measures, get a robust input on the measures, we're going to have you discuss all of the criteria for each of the measures. We're going to have you vote via SurveyMonkey link, which Ann has emailed out to you all.

So, we won't be getting live voting results during this meeting. We'll just have a discussion of each criteria vote and then move onto a continued discussion.

After the meeting, we'll be holding two conference calls for the people who haven't been able to attend, where they'll be able to join with the measure developers, and certainly any of you are welcome to join too, providing them with a meeting summary of this meeting.

MS. WILBON: Can you use your microphone, please?

MEMBER PARE: So, we're going to vote today and they're going to vote later?

MS. TIGHE: Yes.

MEMBER PARE: Why don't we all vote at the same time?

MS. TIGHE: They're not available to participate today, and so --

MEMBER PARE: Yes, but they're going to vote -- why don't we all vote later then?

FEMALE PARTICIPANT: Because we're not going to be on those calls.

MS. TIGHE: Well, if you wanted to join the call and vote later, you potentially could do that. We just thought it would be easier --

MEMBER PARE: It just seems --

MS. WILBON: So, the reason we're doing that is because we tend to have to bug people a lot to submit after. We'd rather just

get people's input while we're here. It just saves a lot of like e-mailing back and forth when people have to vote via SurveyMonkey, it just -- and getting everyone on the call again together, corralling, these space quorum issues again, so --

MEMBER PARE: Is that a valid way to vote though?

MS. WILBON: It's kind of what we're faced with at this point.

MEMBER PARE: Okay.

MS. TIGHE: Yes, so, we'll giving them the benefit of your discussion, informing them via the summaries that we put into our draft report, and then having the opportunity for them to ask questions of the developers.

MEMBER BAYEWITZ: Would there be a way for us still to see live, what our general consensus is here? I'm just thinking about the last meeting, where there was a re-vote after people saw the results.

I wouldn't want to have a situation where we vote and then they vote, and then we find out, "Oh God, we need to do this again."

Because you know, if 75 percent of us are all in agreement, mathematically it won't matter for that one additional person.

MS. WILBON: So, the other thing to keep in mind is that there will be an option to re-vote after commenting. So, given that every -- we have a quorum on that call and there is an opportunity for everyone to participate, depending on how the votes come out, you know, there will be an opportunity to discuss again and vote again, just like you guys did for the Phase 2 measure.

So, we recognize it's not ideal, but we're trying to figure out the best way to accommodate getting everyone's vote and having everyone have an opportunity to participate.

MS. TIGHE: Okay, so certainly, we understand it's less than ideal. We also, to your point, Ariel, we didn't want your votes to

potentially sway those who haven't voted, so that's the reason why we have the SurveyMonkey link, just to kind of maintain that ability to be impartial when voting.

Okay, next slide. So, at this point, if there no more questions or discussion, we are ready to move into the consideration of the first measure.

So, we'll ask NCQA to come join us at the table.

CO-CHAIR ASPLIN: Good morning. Welcome, and we'll hear from the developers first, 1560 relative resource use for people with asthma.

MR. HAMLIN: So, my name is Ben Hamlin. I'm the Director of Performance Measurement at NCQA.

I'm going to discuss overall, kind of the two measures together, because the methodology and the reporting strategies are the same for the resource use components of the measure, addressing two separate conditions.

These measures that NCQA reported along side the HEDIS quality measures for the same domain, for these same populations to get to our value equation, and obviously, the quality measures are slightly different for each of the different domains.

The relative resource use measures at NCQA are total resource use for members who have been identified with a condition. So, it's not either episode based or to be related to the condition itself. It's all services delivered during the measurement period.

We ask plans to use the standardized pricing tables that we provide. So, it's not actual cost. It's standardized costs which helps us get around some of the issues with, you know, price variation and market variations and regional variations that occur, and it=s a bit of a Wild West in that regard. So, we use these measures to compare health plan performance.

These measures are risk-adjusted using the modified HCC's from CMS, and we

provide the plans back their individual expected information in detailed plan reports every year, based on all plan submissions, and currently right now, there are well over 1,100 plans reporting RRU for HEDIS. So, that's a fairly substantial base of information that we use to tackle these thresholds. So, I think I'll just leave it at that.

CO-CHAIR ASPLIN: Very good. Any high-level questions for the developers, before we move forward to the criteria?

Next one, and then I would ask the Committee, as we move through the various domains for endorsement, if we could just try to stay as disciplined as possible, to stay within the domain and not bring in scientific issues necessarily, while we're talking about importance and et cetera. Sometimes, these are open to debate. I certainly understand that, but it would be much more efficient, if we can keep all of our comments in this same domain. We'll get to everything,

as we move through it. Thanks.

So, we'll start with importance to measure and report.

MS. WILBON: So, the first criteria for 1a for importance is around whether or not a -- sorry, whether or not there has been a high-priority area that's been identified, that this measure addresses or whether or not the developers demonstrated that this is a high-impact aspect of care, affects large numbers. There is large variation, in terms of cost of resource use in this area.

CO-CHAIR ASPLIN: Great, and we are going to hand it over to our lead discussants. I'll just call on the lead discussants and again, ask them to stay to the criterion of importance to measure and report, and while we're preparing -- or if anybody has had trouble getting to the SurveyMonkey link, maybe you could raise your cards, so that we can solve that for you, while we're moving through the conversation.

With that said, Nancy or Joe. Joe, you're going to talk for the -- okay, go ahead then.

MEMBER STEPHANSKY: One thing I want to start right out with, given some of the comments, the written comments that we got back, is the misunderstanding about what it means to be a plan level measure, because there seems to be an emphasis upon going back to the providers, because the providers are who the patients, not interact with, but when -- and we've been dealing with accountability at the provider level for a long time, and it's kind of refreshing to me, coming from the hospital area, to run into measures that might eventually develop into a way of defining accountability for health plans.

Again, we do seem to have some misunderstanding about what it means to have a health plan level measure and how those are actually being used in the field.

One thing that I will say is having

dealt with some large insurers, the NCQA measures, both the quality and the cost measures are taken very seriously by the plans that I interact with, and the fact that they're willing to consider these things as credible and actionable, are I think important to the group.

I don't know if NCQA, if either of you would want to add anything to that, about the importance that plans attach to these. I know that Blue Cross/Blue Shield of Michigan, for example, has considered adding a specific NCQA compliance committee.

I don't think they've put that into place, but that tells you something about how serious the health plans take these things, and I know that they have used the data from NCQA in looking at plan-benefit design.

So, there are ways in which this data gets used, that we may not all be aware of.

MR. HAMLIN: So, I have two points to make there.

You know, one is that we do provide the plans, very detailed plans, specific reports that give them their benchmarks for each of the individual service categories for their peer group. So, it does provide the plans kind of -- I don't know, a competitive edge, but it give them competitive information, if you will, to sort of see where their position is at least, for the measurement time frame.

On the other hand, the plans only submit aggregate information to NCQA. So, we don't get the patient level/member level detail.

So, it's hard for us to really identify true quality opportunities, you know, for each individual plan, even despite the fact we have a huge amount of data from each plan for each of these measures.

But you know, we have ideas of where, you know, we can make suggestions, and our plans can do much more sophisticated analysis of their own data, using both the

expected and true-cost data plugged into the same formula to see, you know, their actual position, relative to the benchmarks we provide.

On the other hand, the people that are very interested in these measures are very much on the employer side, in the employer community and the purchaser side, who like this information because it gives them, even in the broad perspective, of some of the value offered for the price of the services that are being provided by each individual plan.

So, you know, while individual quality opportunities may be more difficult than a traditional quality measure, you know, there are many aspects of these measures that people find very useful, depending on who the stakeholder is.

MR. REHM: Can I add to that? Thanks, Ben.

I'm reminded of Larry's comments earlier about the -- I think he brought up the

directional issue, and I think that by -- I mean, this is a measure that's now -- we were just looking at some of the old test data from 2003/2004.

So, the development phase was incredibly long, and in that time, you know, it's clear -- we're trying to send a signal, NCQA tries to send many signals and it has a portfolio of signal senders, if you will. This is just one of them and it's in an area where there is really a dearth of measurement.

We really are appreciative of the fact that this probably isn't where we want to land, in terms of getting true accountability and getting at true cost, because those are -- they're slippery and there is not really yet a consensus by all the stakeholders who are involved in the process, to be as transparent as we would hope.

But that frame is changing and these measures or ones that may supersede or follow it are going to change accordingly.

We're reasonably sensitive to the market and we don't want to have measures that are no longer addressing important areas.

So, we understand the limitations about that, but I think, from a signaling aspect, to not have a measure or measures like this in play, would be to really, from our perspective and our mission perspective, to be not doing our duty.

So, I appreciate all the comments before that, but I just wanted to emphasize that.

CO-CHAIR ASPLIN: Nancy, do you have any other comments to add?

MEMBER GARRETT: Yes, just a couple of things to add to those comments.

I think the question about the level of analysis does come up every time we talk about these measures, and I think it's a good point. It is kind of refreshing to see measures at the health plan level, but at the same time, for actionability, how can this be

constructed at the provider level? Is it drillable, because really, what you're looking at is a collection of providers and what they're doing with their patients.

So, I think that's kind of a question that some of the Committee members raised.

There was also a point raised about asthma, prevalence and treatment is different in children versus adults, pediatric versus adult. Would it make sense to have two measures, this really aggregate set into one concept? Does that make sense conceptually, clinically?

So, I think that's another question that was raised, and then the question about identification of asthmatics came up here, as well, which I think fits into this, you know, in terms of the overall criteria for the measure, some of the concerns about over-identification, which I think that TEP had raised.

So, I'm not sure if this is the right place or not to discuss that, but those are some of the comments we got.

MEMBER STEPHANSKY: Yes, and that last point, I think is quite important in that these same definitions come up in different committees, because those clinical measures are -- in other committees we don't interact with and, as the science changes over time for example, we don't really have any way of coordinating a response across these different measures to get definitions changed, say at one time, or you know, are we, as a Committee now, if we see something wrong, do we have to raise that and therefore, force other committees to re-look at these definitions.

NQF, I don't believe, we don't have a way of coordinating that, do we?

MS. WILBON: You mean broadly, the definition of asthmatics?

MEMBER STEPHANSKY: Or here, because we have a HEDIS measure or measures,

that use the same definition. If we were to challenge that definition here, what is the consequence for those other measures, or is there a way that we would look at the definitions across multiple measures at the same time, if there were a problem? I'm not sure there is, this time around.

MR. HAMLIN: So, I mean, we did participate in the clinical measures in the first round, and the same issue keeps coming up with the asthma definition for persistent asthma through claims, which is why we keep testing it and keep looking at it over the years, to see if there is a better way to do it.

Right now, because we use claims only for these approaches, this is the best definition we can come up with, and it is the most sensitive and specific for persistent asthmatics from health plan claims that we can come up with, and that's been sort of validated over time.

That being said, we recognize there

are some limitations to the definition and there is probably a better way to do it, if we can get more clinical information into the definition for it, and this definition is consistent across all of the HEDIS asthma measures. The same population is identified the same way, using the same claims.

So, we appreciate this feedback. So, we continue to keep pressing and testing and looking for better ways to do it, but right now, this is -- you know, as I think was mentioned earlier, is through claims that we have access to, this is the most specific and sensitive definition for persistent asthmatics that we can get. It's not perfect.

But we do appreciate the feedback because we do take that back and then build that into our development process.

MEMBER STEPHANSKY: And I have some confidence in the NCQA process, after having followed some measures through your organization, but I'm not sure that we can --

I could have the same confidence in all measure developers to keep looking at it. So, that's why I'm asking about an NQF process for this.

MR. HAMLIN: Well, thank you for that.

DR. BURSTIN: So, there are a couple different responses to that.

First, you know, as part of when asthma measures come up, and these guys lived through this recently, the Committee does, in fact, take a pretty close look at the definitions of the populations.

Again, it's often complicated by different data sources and issues along those lines, but they try to get the numerator and denominator, at least in terms of the science and the population, as close as they can, and you know, an important point that you said several times was, given claims data, and I think the real issue is now, moving and thinking prospectively about what you could do with eMeasures. There is a whole effort now, as

part of this to create value sets and work on harmonizing those value sets and NCQA has been doing a lot of that work, as well.

So, I think some of that is coming, but I think it's kind of the best we can do with the data sources we have. That's part of the problem.

CO-CHAIR ASPLIN: Andrea, any comments from the TEP on importance to measure and report?

MS. WILBON: So, the TEP didn't really evaluate importance, per se. It was more in the scientific acceptability portion.

CO-CHAIR ASPLIN: Yes, okay.

MS. WILBON: So, we can --

CO-CHAIR ASPLIN: Then open it up to the full Committee.

MEMBER GELZER: I've got to get to my notes, please.

Okay, so you want the TEP --

CO-CHAIR ASPLIN: If your comment -- if your comments from the TEP perspective,

I only was just going to you, to stay on this criterion.

So, if you have --

MEMBER GELZER: Well, I actually have a comment to -- I feel I need to respond to a comment that was made here.

CO-CHAIR ASPLIN: Great. Why don't you do that, and then we'll open it up to the Committee.

MEMBER GELZER: Okay, very good. Just a point of clarification. Health plans deliver population-based strategies.

We manage costs and quality across the continuum. We are, we have been, we continue to be held accountable for both costs and quality.

NCQA developed the first HEDIS data set in the 1990s. So, we've been measured on quality metrics and now, they're adding cost metrics, and I think that's a natural evolution, and again, I don't think they're perfect, but I think they're certainly very

well-constructed.

CO-CHAIR ASPLIN: Very good. Cheryl, do you have a comment?

MEMBER DAMBERG: Yes, and tell me if I'm straying into the next section, because I wanted to talk about the measure intent a little bit. So, am I straying?

CO-CHAIR LATTS: That's okay.

MEMBER DAMBERG: That's okay? So, I had a question for NCQA because one of the things that I was challenged with, if the goal here is to marry costs and quality metrics to create some type of value metric, it seems that the quality measures are very narrowly specified around asthma care, but yet this measure, as constructed, looks more broadly across all the care that an asthmatic gets in a year.

And I sort of found somewhat of a mis-alignment there, and what I wasn't quite sure, because I didn't see any data -- if it was there, I'm sorry if I missed it -- related to

what proportion of the total costs for a given asthmatic are asthma specific in a given year, because I think that there is sort of lack of specificity in this measure.

MR. HAMLIN: Yes, so, that's a great point. You know, the strength of the value equation is dependent upon both sides of the equation and the asthma quality side is certainly not as detailed, because we use the existing measures that are available and HEDIS said are only allowed for public reporting.

We've had additional conversations about other services that could be included in that quality composite, some of the prevention and screening measures for immunizations and things that might be appropriate for the asthma population, that could be informative of that quality dimension.

There are other additional measures that are being developed for, you know, patient-reported outcomes and patient assessments for asthmatics, and people like

that, that I think would be very important to have in this, and unfortunately, until those a fully developed, tested and available, you know, for public reporting status, which I think is still a little time away because of data issues, really and truly, the quality dimension for the asthma RRU is still quite limited.

However, we think it's critical to include it, and so, therefore we do include it as part of our value equation.

Unfortunately, that's probably not the answer you wanted to hear. We do not -- we cannot look specifically because again, we get aggregate information on the different patient cohorts for the different populations at RRU. We cannot split out the episode -- I'm sorry, the condition specific treatment versus the non-condition specific treatment, and then again, that was kind of the decision that was made early on in the development, in discussion with plans and the committees, such as this one,

because that is a rabbit hole that you continue to spiral down, and you can get into conversations about what is attributable and what is not.

So, just looking at the total resources for health plan use per year, to manage populations, gives us a comparison model for other plans who are managing, and assuming that using the risk adjustment methodology, the management strategies and the patient severities and things are relatively comparable, using a fairly specific risk-adjustment model, that the results can still be compared to -- you know, within the peer groups.

MEMBER DAMBERG: So, is there a reason you didn't proceed with an episode based approach?

MR. HAMLIN: There is, because the idea of, you know, episodes attributed to the condition became sort of endless debate.

You know, the examples that I use

are, you know, asthmatic attack and like, child fell of the monkey-bars and broke their arm. Well, is that services used to treat the broken arm attributable to the asthma, because it was an asthma attack, but that may or may not show up in the record again.

So, total services per the measurement period for someone identified with a disease was a more reliable method for us to be able to measure and compare plan performance, you know, within their peer groups.

CO-CHAIR ASPLIN: Mary Ann?

MEMBER CLARK: Yes, thanks. I just wanted to say that, you know, I think there is no question on the importance to measure and report. You know, that's obviously a prevalent, costly condition, and at the plan level, the measure -- I mean, it seems to me that it is appropriate for measuring plan-to-plan performance and at a high level, you know, plans can compare themselves to each other. I think

that's the goal, right?

Then they are -- would be able to then determine where they need to focus their efforts.

I guess one of the questions I had in this section had to do with reporting information over time, and I know that this has been, I guess an issue because of the ability to compare how plans have changed over time, because the measure is actually comparing to an average in a peer group or other types of comparisons.

I know that you've been working on that and I was just curious how that is progressing, because in some of the -- for example, I know in some of the CMS measures, for example, they are looking at two different tiers of performance in some of their measures, like plan -- or improvement of the institution, for example, if it's at the institution level to themselves, and then improvement to -- over time -- you know, as their peer group.

So, I was just curious on how that is progressing.

MR. HAMLIN: Right, so, these are again, conversations we continually have.

The benchmarks for the RRU measures are basically calculated using all submissions for that year, and so, in order for us to compare year over year, we'd have to hold a number of things constant.

First of all, the standardized prices. So, we'd have to basically freeze prices at a certain level. We'd also have to freeze a number of plan submissions at a certain level.

So, we would probably lose a great proportion of the plans because of the way plan ID's are used to identify plans that are reporting to HEDIS. You would probably see a huge drop-off.

So, there would be a few plans that perhaps, could compare their performance year over year over year, but again, we're still

looking for the value in that comparison because again, you know, everything being normalized to the average for each plan on a yearly basis, in sort of a snap-shot fashion has been what people find useful, as far as where things shift.

We do look at the -- you know, as I think you saw in the testing report, the quadrant shift, to see if a plan significantly, so that suddenly, something happens and something significantly changed, but that's really been the only indicator that people have really found useful, I think given that we would have to hold so many other things constant, the final result of plan tracking through the quality and cost dimensions probably would be too artificial, I think even really, because we would have to unwind so many of the different calculations that go into the actual RRU calculation to begin with.

So, yes, we're still having this conversation about, you know, is that possible

and how can we provide less static reporting methods, to try and show maybe confidence interval and things around plan results, to give at least a much more meaningful information, but we haven't figured out the perfect method to really report that tracking of quality. The tracking of resource use changes over time for each of the specific categories, or even at the aggregate level, but at the total medical level. We haven't given up yet though. We're still trying.

CO-CHAIR ASPLIN: All right, thank you for the conversation. Ariel?

MEMBER BAYEWITZ: Yes, I would just echo what Joe said before that, you know, from a plan perspective, we take these measures very seriously. There is -- I mean, I could speak for, you know, WellPoint. There is very large teams that look at the NCQA, specifically HEDIS measures.

We have proactive programs to reach out to providers. We have -- you know, I think

of one program which has more than, you know, around $15 billion in spending that's tying to these kind of measures.

So, we definitely take them very seriously. I think the question is, in terms of resource use, you know, what is the value of looking at that at a plan level, and I think from an overall cost perspective, to what Andrea mentioned, I think there is value.

I think there are things that a plan can do to drive cost. I think about our referential pricing programs, which have significantly reduced certain costs in certain areas, or bundle payments.

I mean, there is a lot of things that you can do, but a lot of that is intimately tied to unit cost, and I don't know if stripping out unit cost from this equation gets you to a place where looking at a plan comparison is meaningful.

You know, I think the next step is always going to be, to Nancy's point, okay,

well, who controls this, these resources, and that's the provider, and if at the provider level, you can't evaluate the measure, then what is the point of rolling it up at the plan level?

You know, unless -- and maybe we'll speak in a bit, we can prove that provider roll-ups are meaningful, I don't know if, you know, if this is going to make significant contributions to understanding, you know, cost at a plan level.

MR. HAMLIN: I think our perspective on that has been primarily that if anyone has a great lever to help encourage providers to change habits or to, you know, change utilization patterns, the plan is probably the first target, if you will, and I'll say target.

But has probably the -- you know, can get both hands on the lever and try and change those behaviors or, you know, through their reimbursement policies, through their

payments, I mean, you mentioned bundle payment, I mean, I was just at a bundle payment seminar last week in the Midwest and they were talking about, you know, the -- because of the nature of bundle payments, you know, you really work within very strict confines of, you know, what you're getting reimbursed for.

So, therefore, it's really important to drive quality and efficiency because if you don't, you end up having to pay back money or you end up having to, you know, lose out on significant amounts of financing.

So, again, you know, thinking of it in that context and by providing this measure that gives you very detailed information, there is 37 different categories of information provided back to the plan, you know, plan-specific benchmarks that are calculated from their peer groups, their peer group at a national and regional level, to really try and show them their positioning and where those opportunities may lie, to further dive in, and

we think that the plans really --there has been some limited testing in provider groups using these in -- using the structure for the measures, which has been very interesting for the plans that -- at California.

You know, limited number of plans that operate in that space, and so, they were able to really look at that provider group information, but again, you know, it requires those benchmarks and those expecteds from the average peer group performance at these very specific levels, and to do that, you need big data sets.

And so, that is -- so, our perspective has been we provide it at the plan level, we provide you the benchmark specific to the plan, the plan at the national perspective and the plan at the regional perspective, and it's HHS region, in the hope that it gives the plan enough information that they can then go and use special programs to try and drill down into that data, using other specific data or

specific program.

You know, there are HEDIS shops, if you will, to try and identify, you know, places where they can really maximize their efficiency through the quality services being provided, without losing the quality, obviously, because that's most important.

CO-CHAIR ASPLIN: Lisa?

CO-CHAIR LATTS: So, my comment to that, Ariel, is that that's what we used to say about quality, you know, in the -- back in the early days, the late 90's and early 2000's when all the HEDIS measures started coming out, we used to say, "Oh, you can't hold the plans responsible for quality. That's not our problem. That's the provider," and we don't even talk like that anymore, because it's just accepted that the plans will do -- take efforts to improve the quality of care that their members receive, and I think cost will get there, as well, that it will be just taken for granted.

My sort of issue with it is that this is a measure where you don't want to be too high and you don't want to be too low, because even though we talk about efficiency, and sort of the cost quality equation, with the limited quality measures we have, you can't feel comfortable that a very low cost with the limited quality measures we have actually is in a good place.

So, really, where you want to be is clustered around one and clustered in with everybody else.

So, that's sort of my issue with these measures, is that it's not really getting to value because you just want to -- you don't want to stand out from the pack. You want to be in the middle.

MR. HAMLIN: I don't know if I really -- I think I might argue with that, in fact.

I mean, maybe for the total medical roll-up you might want to be clustered around your peer group, because it makes sense.

We've seen, even with the limited quality information, we've seen huge variation in plan resource use at the same level of quality, as the early -- I think you may have seen in the early days.

But I think the message here is really, we don't -- we've very specific that we don't say higher or lower is better for either -- well, for resource use. Certainly, quality, we are very much in one side of that equation.

But you know, I would argue that, you know, a plan being higher than average in their outpatient E/M resource use, and that's driving very high quality scores, would probably be a very good thing, and if it's higher than their peers, it means they're investing in those services to get patients into that outpatient E/M, and I think that's a good use of their investment.

I don't judge them on that, but I think that doesn't -- that doesn't -- that makes

sense to me, if you will, and those are the -- I think those are the things, but by again, splitting the measures out into specific service categories, I think there are some things --

I mean, you know, obviously high ED utilization is bad. No matter what, there is no argument for that, that it's ever a good thing.

But on the other hand, we don't say higher is better, lower is better for these measures, very carefully, and we're very deliberate about that, because I think certain high services for high quality is good, as long as the quality is high, even if it's limited, you know, and we're always hoping to add more to that quality equation to help, you know, again, as we did with asthma.

We just expanded to now, three measures in the composite, which gives much more variation on the quality dimension.

The resource dimension still kind

of looks fairly similar, still broadly distributed for every plan, you know, quality level, but you know.

So, looking for high quality reasonable resource use clusters in those roll-ups, I think is probably where you want to be, not necessarily just at the average, because average is just average.

CO-CHAIR ASPLIN: Bob, and I think we'll move on to a vote.

MR. REHM: It will be a really quick comment.

I love to disagree with Ben, and I also like to agree with Lisa.

You know, one of the conundrums on cost, and you know, the most typical example is, especially for measures like this, is pharmacy costs, which we do capture in the measure, and in many, many ways, you would want to see a higher utilization, better adherence, appropriate medications applied, you know, regardless of where it may reside on the

formulary, and that's going to -- you know, in the year or whatever, that could increase your cost, maybe downstream of course, the ROI on that is very attractive and a lot of people like Mark Fendrick, would suggest value-based, you know, benefit design would make that a smart move.

So, you know, in some ways, you're right, where is the right place to be on that bubble? It's really up to really the community, to figure that out, I think.

CO-CHAIR ASPLIN: Very good. Thank you for the conversation.

So, we are going to move into the vote on the criteria, first the sub-criteria and then the overall importance to measure and report. Ashlie, do you want to walk us through?

MS. WILBON: Sure. So, hopefully everyone has been able to open their SurveyMonkey link. No? Okay, I think we have -- do we have paper?

So, we have some -- we came prepared, just in case we had IT issues. So, we have a paper version. If you just want to circle what your votes are, we'll enter them on the back end for you.

So, if anyone else is having IT issues, just let us know. You can just keep track of that and just give it to us at the end of the day.

What I'll do is, why don't we just walk through all the importance criteria, and you guys can just take a minute or two, to just think about your votes and we'll continue on and start the scientific acceptability criterion.

So, again, the first criteria within importance to measure and report is around the high priority, which I've already discussed that they've demonstrated that there is variation in performance in this area, and that it's a high aspect -- high-impact aspect of healthcare.

Opportunity for improvement, that

they've demonstrated that there is an opportunity, again, variation with an opportunity for improvement or a gap in performance, such that there is room for providers or health plans to improve on the measure.

Then the last one is around measure intent, and whether or not the intent is clearly described and that the resource categories that are listed support the intent of the measure.

Then the last vote again, is on the overall importance. So, I'll give you guys -- considering all those three components, I'll give you guys just two minutes or a minute or so, to enter your votes on SurveyMonkey, and you can just keep it open, as we go for the day and we'll go from there. Let us know if you're having any technical difficulties.

I would just keep it open. Can you just keep the survey open?

MEMBER GELZER: Keep the survey?

MS. WILBON: Open.

MEMBER GELZER: Okay.

MS. WILBON: Yes, so, don't -- yes, don't close it. Just keep it open for the day. Hopefully, it won't time out. If we have those issues, we can deal with them.

CO-CHAIR ASPLIN: We are going to take a stab at going all the way through scientific acceptability and then we'll take a break, okay?

CO-CHAIR LATTS: Figured it would speed people up, if you didn't get a break until after it was over.

CO-CHAIR ASPLIN: So, if folks have completed their voting, I think we can move on to scientific acceptability and Joe, did you want to lead the way again?

MEMBER STEPHANSKY: Given the data source, let's see, oh, the specifications on reliability.

I didn't really have any issues with the way that they presented their reliability approach. So, I am not in a real good position

to offer any criticism on this particular area.

I do have some concern that some of what we are seeing in the measure still comes from problems at the plans, essentially from their IT and coding areas, where they're preparing the data to send to you.

My own experience is that in insurance companies of any size, these responsibilities get put in different silos and some silos are very good in their analytical and data processing capabilities, and other silos are not so good, and that I know that you are trying to do auditing, and I guess I would be somewhat interested in what kind of audit problems you have run in to with this.

MR. HAMLIN: So, the RRU measures, because of their immense size, if you will, I mean, there's almost 70,000 data points per measure that get submitted on occasion to NCQA, which pretty much over -- goes well beyond the entire HEDIS set for quality care.

So, we've been pushing out for the

last couple of years, the measures in electronic form in XML, to provide electronic interfaces for plans to -- once they've got their systems up and aligned and their coding up and aligned, to be able to report the measures more accurately, sort of reducing the human error, if you will, on the input side.

The auditors, in conjunction with that, we have NCQA certified auditors that must sign off on all the data before it gets submitted to NCQA for HEDIS reporting, have developed very sophisticated tools to go in and use those electronics and that coding to validate the data sources and the, you know, primary source verification and things like that. Much of this is becoming really very technology driven.

So, we haven't heard a lot of -- you know, lately, because the measures have remained fairly stable over the last few years, a lot of really interesting auditor comments, like we do for some of the more clinically

focused measures, because again, we've really tried to remove some of that error component as much as possible.

You know, again, do errors probably exist because of the volume of data that is required to report the measures? True, yes, I'm sure there probably is some, but again, we do everything we can to maintain the measures as standard as possible, the cost and the standardized pricing tables as much as possible, and like I said, by pushing this out in XML to plans as well, in a very detailed fashion and keeping that very consistent, you know, it should theoretically be easier for plans to update their systems without having to re-code the entire measure every single year for reporting purposes, and we think that is a great leap forward in reducing error, as far as reporting the measures.

MEMBER STEPHANSKY: Theoretically.

MR. HAMLIN: Theoretically. Most

of the plans we deal with are extremely sophisticated and also, NCQA also certifies software vendors to do some of these plan calculations, in addition.

So, we use extensively Test X, with thousands and thousands of patients, to validate software certification, or the software vendors who are calculating these results, as well.

So, there is many ways that we have confidence in our data, if you will, through these different programs that we use for each report.

CO-CHAIR ASPLIN: So, other comments on reliability? Nancy?

MEMBER GARRETT: Well, I'll just call out a few other comments that the Committee had made, through the process here.

On risk-adjustment, there were some mixed comments. Some people thought the risk-adjustment model seemed to be working pretty well, but other people thought the

R-squared was low, and had some concerns, and then the issue with socio-economic risk-adjustment came up pretty strongly, saying for asthma, it's like we have a big effect, and so, that was a concern.

So, one thing for the committee to be aware of is that while the current guidance from NQF recommends -- basically says that you can't use these socio-demographic variables in the risk-adjustment model, you can stratify, you can recommend that the measure be stratified after the fact, and that you use it by stratifying by the groups that you're concerned about, in terms of the impact on the measure.

So, that's something that we could recommend. So, keep that in mind, as we talk about it.

Then some people said that they didn't know if they had enough information specifically on validity and reliability for this measure. So, I don't know from the TEP,

if you want to add some things.

MEMBER GELZER: Yes, from the TEP perspective, I don't think there were any show-stoppers, as far as reliability was concerned.

They did make the same comments about risk-adjustment and that referral centers would perhaps, be adversely affected.

On the follow -- one of the follow up calls, I don't know if it was -- I think it was the call, the joint call with the Committee, with our Committee, that NCQA responded that they were closely following NQF's work on stratification and that -- just from my observation of the discussion, appeared to alleviate concerns.

CO-CHAIR ASPLIN: Andrea, was the TEP satisfied overall with the diagnostic criteria for inclusion in the measure, because that seemed to be one of the topics that they were most interested in and in our conversations together with the developer and

phone calls, there was still some questions.

However, I think your characterization of -- my take-away was the same, that it wasn't a show-stopper, although there was questions about single ED visits, etcetera.

MEMBER GELZER: So, the inclusion criteria were a big topic of discussion, as well as the measurement time period, and the TEP was concerned that -- I think certain individuals on the TEP were concerned that you could get an asthma diagnosis if you went into an emergency room one year, and it's a two-year measurement period.

I believe that NCQA clarified that you would have to have a diagnosis in each of the consecutive measurement years, and once that clarification was made, I think that the TEP was again, appeased.

MR. HAMLIN: Yes, so, it's a two year identification period. It's only a one year measurement period.

So, that is true, you need to meet the criteria in both years of a two year identification period, in order to get in, and that was, I think -- that clarification was what the -- helped the TEP come to --

CO-CHAIR ASPLIN: Other comments or questions on reliability from the Committee?

MR. HAMLIN: Did you want me to address the issue of --

CO-CHAIR ASPLIN: Sure.

MR. HAMLIN: So, the measures themselves are reported by product line, and for asthma in particular, has the highest number of age stratifications of any measure, and that's designed primarily because of both clinical factors, the younger age population is separated out, but also reporting programs.

So, there is a stratum there for, you know, child health programs which generally, the cut-off is around 18, the Medicare -- Medicare is only -- sorry, Medicaid is only reported with Medicaid plans.

Commercial is only reported with commercial plans, for each of these individual age strata.

So, while we are very closely following the SES conversations, and I'm probably providing more feedback than NQF staff want to hear on the SES risk-adjustment, you know, these -- through stratification and through using product line reporting comparisons, we feel we're as close as we can come at this point in time, again, without further specification on what's appropriate for SES reporting.

CO-CHAIR ASPLIN: Okay, I think prior to the vote, we want to walk through the algorithm that's in front of you on reliability. Do you want to?

MS. WILBON: Yes, well, I --

CO-CHAIR ASPLIN: It's at a very high level.

MS. WILBON: I will just say that we're not going to strictly enforce it, but it might be good to just highlight some of the

questions you should be considering or how you should be framing your vote on reliability, in particular.

So, the questions around reliability are around whether or not there was adequate testing or empirical testing, whether or not that testing was at the measure score, the data element level or both.

Your vote based on the scope of testing of high, moderate or low is based on your confidence on whether or not, you know, the measure score, the testing demonstrates that the measure score is highly reliable, moderately reliable or not reliable, and the same for the data element, depending on which route they took, and I'm pulling up their testing and I believe they did testing at the measure score level, is that correct? Yes, so, okay, okay.

MEMBER RATLIFF: I don't want to keep us away from our break, okay, so, quick question, but let me keep us away from our

break.

So, you identify asthmatics based on a claim with a diagnosis of asthma? Do you see a lot of patients drop out over that two year period, where maybe they have a claim with asthma and then they don't have any further claims?

With adoption of EHR's by clinicians, it seems that once that ICD-9 for asthma goes in, it's going to perpetuate for ever.

So, have you seen a change in those drop-outs or a change in that reporting over the --

MR. HAMLIN: Well, so, ICD-10 was supposed to solve everything, because ICD-10 is specific to just asthma, and we all know where that is, at this point in time.

So, we are looking at dual coding practices to see how the plans are actually converting the clinical data to an ICD-9 code for persistent asthma -- well, for asthma, to

ICD-9 for ICD-10 for persistent asthma. So, to see how that all affected our denominator potentially.

On the e-measure space, we're really designing algorithms that look really for persistent asthma in the problem list at the current moment. I personally feel that's not probably specific enough for what we're looking for.

So, I'm continuing to push, sort of how those algorithms are done, but that work has really gotten down into some details and they've been bogged down, as far as what we can use to identify patients, which is why it's still in e-measure theoretical development space and not really sort of in HEDIS programs.

You know, the HEDIS algorithm right now using ICD-9 requires a combination of multiple encounters with a diagnosis of asthma, with medication events, medication events alone under certain criteria, and over a two year period.

So, you have to have -- meet those criteria in both years, and so, by doing that, we feel we do weed out most of the seasonals and the non-persistent asthmatics. You know, again, the sensitivity of that identification algorithm gets better, the better the data gets and the more specific the codes get.

So, we're hoping to see clinicians coding persistent asthma in the future and in the e-measure space coding -- you know, again, identifying persistent asthma versus just writing a generic asthma diagnosis in the claims and requiring for us to take a look for additional elements that might help define whether that patient is truly a persistent asthmatic or not, which probably again, will look for multiple encounters in the healthcare system over time, or medication events, you know, consistently throughout the measurement period, to try and refine that algorithm, but those are still very much in development and testing at this point, and testing has just

begun on those, to try and see how they compare to the current algorithms.

CO-CHAIR ASPLIN: Any other questions prior to voting on reliability, as you review the algorithm?

MS. WILBON: So, again, questions to consider as you're voting, whether or not reliability testing was conducted, whether or not the method was described, the approach that they used to test reliability, such as signal-to-noise, random split-half correlations and so forth, and because they did both data element and measure score reliability, you'll want to ask yourself questions around both of those.

Then again, your scoring or your rating of high, moderate, low is based on your certainty, your confidence level on whether or not the measure is reliable, high, moderate or low.

Does anyone have questions about that?

The other piece of reliability including -- in addition to the testing is around whether or not the specifications were clear and precise.

So, if you want -- if you guys are ready to vote, we can take -- pause for a minute or two, to let you do that.

CO-CHAIR ASPLIN: All right, let's go ahead and complete the reliability section. Thanks.

So, we've collapsed this, just to one vote on reliability.

MS. WILBON: Yes, there is only one, yes one vote. The other elements were just kind of a reminder of what's included in the specification.

So, if we're ready, just a quick reminder.

Generally, because of the -- if you guys remember from the last meeting around the must-pass criteria, because we're getting the votes in pieces from the Committee, we're just

going to go through all the criteria for all the measures, and we'll see where we land at the end, once we've had all the Committee weigh in.

So, we're not going to stop. Obviously, we're not going to stop and calculate or anything like that. We'll just continue to evaluate.

CO-CHAIR ASPLIN: Thank you. We'll stay with the same order. Joe, any comments on the validity?

MEMBER STEPHANSKY: On validity, I think we'll -- as we discussed at our last meeting, much of what we are looking at is face validity, and I found that was really pretty good, from my standpoint.

The one area where we had people making comments was perhaps on our -- the risk-adjustment approach, where we were using a modification of the CMS hierarchical condition categories, yes, and my use of that, I have found it to at least go down fairly low on the age scale and still be appropriate, but

others are saying it hasn't been appropriately validated for the non-Medicare population.

So, is there any discussion we need to make on that?

CO-CHAIR ASPLIN: I would lean on our methodological experts in the room. We have four others, but Andrea, do you have a comment?

MEMBER GELZER: Yes, I'm not a methodological expert, but one would think, just from a common sense perspective, that the younger asthmatic diagnosis would be more reliable, wouldn't they? More valid, because as you get older, then you have COPD and other stuff complicating. That's just --

MR. HAMLIN: Right, so, two points to that. When we tested the appropriateness of the different risk-adjustment models, it was not -- before I came to NCQA many, many years ago.

They looked at four different models for appropriateness of the measure.

They did do some additional testing of the appropriateness of the HCC to the broader population and across asthma and COPD. So, they'd look at the younger.

I can't speak to how sensitive it was down at the five to 11, you know, asthmatics, if you will, but there were differences in the different populations, but they were not deemed to be significant enough to make the model invalid.

You know, to your second point, you know, again, not having an extensive testing data in the HCC down at that age level, it's harder to sort of say whether it would be, you know, more or less valid, but you know, despite the fact there were differences noticed in the testing, they really were not -- they didn't -- weren't significant enough for us to be concerned about them, and so, we felt again, for the -- you know, the large populations that were going to be included in the measure, that they were -- the HCC was adequate enough to predict

the utilization, to adjust for the utilization expectations for these populations.

MEMBER STEPHANSKY: And then there were some issues raised regarding the costing approach at different times, not so much in the comments from our members this time, but in earlier meetings.

We are stuck with, for a variety of reasons, having to do some sort of costing approach. We fight this all the time, any time we're dealing with an insurer, because of the differences in contracts that, for example, our hospitals have with the individual insurers, and the necessity for keeping some of that confidential.

I find if there was a costing approach out there, I wish more insurance companies would adopt it. It would be the one that NCQA uses.

So, again, I'm not in a position to be able to offer any criticisms of it, but there may be some out there. Comments?

MEMBER GELZER: The TEP noted that -- but again, that was not a show-stopper. I mean, they just noted that the measure was not specified for an episode of care. It was for a cost of care, and that the clinical -- you know, clinical diagnoses were less considered than the cost.

MR. HAMLIN: Yes, we wish we could publish a standardized cost for every service that's out there. But again, we're really limited to what's available in data, and we again, are very conservative in pricing out the services that we feel we can reliably cost.

So, the things that are included in our standard pricing table are tested every year for sort of reliability, if you will, of that code being used in the huge data set that we use to provide those standardized costs.

So, you know, some of the more rare codes will not appear in the standard pricing tables for reasons that we just can't -- we don't feel like we can reliably price them for

this particular model, and there are adjustments to those -- the standard pricing methodology for the RRU model that are, you know, based on a number of large data sets, but also, sort of for the -- you know, the outpatient versus inpatient.

There is an adjustment there, built into the model too, to account for the resources that are identified as used for that particular procedure or for that encounter.

CO-CHAIR ASPLIN: Nancy, do you have any other comments, and I know Cheryl is on the phone, but we'll have Nancy and then Andrea, if there are other TEP comments first.

Any other comments from the TEP?

MEMBER GELZER: No, I think we've covered them.

CO-CHAIR ASPLIN: Okay, Cheryl?

MEMBER DAMBERG: Thank you. I had a question for NCQA related to the SES issue, and I was curious whether you had run any kind of sensitivity analyses, including some type of

SES adjustment to see what kind of impact that would have, because I know that is sort of this niggling concern about potential validity of the measure.

MR. HAMLIN: So, specific to RRU, we have not run an SES analysis because again, the data that we use is, you know, a large data set from plans, and it's not universal.

We have done -- you know, again, the SES really comes down to an issue of data availability, particularly with certain factors, even at the zip code level, some of that data is buried within the confidentiality contracts of the plans.

So, some plans have very good race ethnicity and SES data available in their data sets, where other plans have blanks, blanks, blanks and blanks 100 percent across all of their data sets.

So, again, it's a matter of, you know, when health plans consistently collect and make that data available in these -- you

know, for us to be able to test the appropriate approach to SES stratification beyond what we already do, as far as age and product line.

You know, we probably will start running those analyses, but right now, it's just too -- we keep testing, you know, the availability of this data within plans and every time, it gets from zero to 100, depending on what you're looking at.

MEMBER DAMBERG: Well, related to that, do you have any sense that the SES mix varies a lot across the plans that you're measuring, because I think if it was randomly distributed, you wouldn't necessarily care, but if there are concentrations of those populations in some subset of plans, that, I think is when you would, you know, be more concerned.

MR. HAMLIN: Yes, and I think this is -- the limitations of reporting this are already at the national level, certainly, and even at the regional level.

I think the further you drill down, the larger those differences potentially could create problems in the measurement strategy.

But because it really is so high level, because of the inherent variations in the -- you know, the way the plan reporting is structured, and you know, plans have gotten bigger and bigger and have covered areas that don't really -- aren't really geographically bound necessarily, you know, state lines are possible still in the data sets, but they, you know, create complexities, once you start getting down to HSA, and who -- which members are from which plan and which HSA or things along those lines, you know, at a much lower level, it gets really, really complicated.

So, again, the concern there is the amount of error you introduce, just trying to get down to that level of detail, and so, we continue to investigate it, but I think at this point, we're uncomfortable with the level of

area that would be in the data, if you will, or the level of unreliability of the data, because of the complexity involved, and just to -- in assigning members and -- or plan responsibility, if you will, to the certain members within certain geographic areas or certain markets.

CO-CHAIR ASPLIN: Could you clarify the R-squared and the risk-adjustment in this space, because there was some comments that were at a much lower number than kind of the one on the third measure, was kind of -- we'll discuss.

I thought it was relatively high on the HCC RRU model.

MR. HAMLIN: Right, so, when we were -- when we did the R-squared testing, initially it was to compare the four different risk-adjustment approaches, and the HCC R-squared in comparison to the other three approaches that were, you know, under consideration was basically deemed to still be

acceptable.

I don't think it was necessarily the lowest, but again, I think there were other factors that were taken into consideration.

Initially, these measures were released with just an age/gender cohort yes/no risk-adjustment approach. It was one -- I think it was model one that was tested, and it was -- because the complexity of HCC, you know, they went on to delay that implementation.

But again, it really wasn't the -- you know, we didn't validate the HCC approach itself because that was -- you know, the model was maintained and developed by CMS, but you know, in terms of appropriateness for this population and perhaps, because the R-squared was higher because we did actually expand the -- the HCC beyond just the Medicare population when you look at the testing, those R-squared might be slightly higher than you would have expected, but it wasn't concerning in the context of comparison to the other three models

that we were comparing it against.

CO-CHAIR ASPLIN: And it's the HCC RRU that is being used in the --

MR. HAMLIN: It is and it's just -- we just don't use all of the CC's from CMS, really so, it's HCC -- it's just -- it's narrowed slightly, not much, but slightly.

MR. REHM: And just an additional point. When I happen to be, at that time, AHIP's representative on the CPM, which voted on -- during the period where the shift was made, and it was really considered a significant improvement on the measure and one that was hard-won, if you will, because of all the concerns.

Historically, I'm going way back in time now, about getting the measure, it's like a stepwise function.

So, the public comment was really quite positive around that.

CO-CHAIR ASPLIN: Nancy?

MEMBER GARRETT: I just wanted to

follow up a bit on the comments about data availability, related to the socio-demographics.

So, certainly, I'm empathetic to that. It's come up a lot on the risk-adjustment Panel as well, and one of the things that we really talked about is if we don't change the way we're approaching this and change the overall recommendation for the need for that adjustment when it -- when there is a conceptual reason to do so, then we're not going to change and we're not going to actually start getting the data.

So, you know, continue to watch that, and that will be happening here in parallel, but I'm hopeful that some change in the NQF approach will also start to change the way the data is collected and start to change the measurement, and so, hopefully, that will all be happening.

But also there is -- I mean, plans have member addresses and you can geo-code down

to the census block level, and there is some increasingly more sophisticated methods around using that as proxies for SES.

So, you know, that is another option that you could potentially explore.

MR. HAMLIN: So, yes, I mean, my mantra is, there is nothing like a quality measure to change the way data is being collected and report -- you know, plans, because when -- once you give them a template to build to, they tend to build it and I agree with you, there are very sophisticated geo-coding analyses that are -- that some plans are doing, and again, once NQF outlines a systematic -- a standardized criteria for, you know, how SES risk adjustment should be performed, we were certainly going to start testing our RRU against that criteria, to provide that standard template for people to build their systems to, to start reporting again, and I am absolutely hopeful that that will drive further collection and more

sophisticated, you know, standardization of the geo-coding analysis, that we can then have a better RRU measure that is stratified by SES.

MEMBER GARRETT: Well, I don't want to over-promise here. So, the risk-adjustment committee is really focused on the question of whether to do this socio-demographic risk adjustment and not as much about the 'how'.

So, the specifics of how this is going to happen is going to have to evolve as the science of how to do a well-evolved, so just to set expectations for what is going to happen.

CO-CHAIR ASPLIN: Bob?

MR. REHM: Yes, I mean, I would -- if you think the measure development enterprise is getting smaller and smaller because of the resource uses required, I think there is a new industry, but I don't know who will fund it, trying to get at the science of risk adjustment and appropriateness for -- I mean, a plethora of measures that are different sizes, shapes and where they fit into the whole spectrum.

So, maybe people at RAND and other folks, the Yale folks and others will help us along.

CO-CHAIR ASPLIN: Both require data, but to the extent that we use a stratification rather than adjustment approach, you may bypass some of that complexity. That doesn't -- your point is still quite valid, about the resources involved in creating these measures.

Gene, do you have a comment?

MEMBER NELSON: It's a question on page 41, the table four, the R-squares for medical plus drug costs are .50 or for medical costs, .48.

At what level do you -- are you concerned about over-adjustment?

MR. HAMLIN: That's kind of a loaded question, actually.

Yes, I'm not sure I can really answer that question.

I think that the -- you know, again,

when we were looking at the R-squares for the different risk adjustment models, it was really in comparison to each other.

Once we had selected the model, we really didn't focus so much on the specificity of the model beyond the appropriateness for this type of measurement approach.

I'll have to get back to you on the, you know, if we actually really have a criteria for threshold, of what is over-adjustment versus what is under, and where the target is being missed.

It's kind of a wide margin at this point in time, just because of the -- you know, the data issues, again, primarily, and you know.

CO-CHAIR ASPLIN: Very good. Once again, if we could do a high level overview of the algorithm, as we prepare a vote on validity, unless there are other comments from the Committee.

Ashlie, if you have any comments on

the algorithm, that would be great.

MS. WILBON: Sure. So, the algorithm, again, highlights some key questions you can be asking yourself as you're jotting down your vote or answering your vote on SurveyMonkey, whether or not, similar to what we have on the slide here, but whether or not potential threats to validity were empirically assessed, whether or not empirical validity testing was conducted, and on this measure again, they did face validity testing.

So, the question is whether or not the face validity testing was systematically assessed and the results demonstrated that it was actually valid.

So, I think those are some of the key questions to be asking yourself and again, there may be some other questions along the box plots but I think the key thing here again is that, remember face validity is the minimum threshold, and if you guys feel that that was adequately addressed within what -- the

materials they submitted, then that should be considered.

Validity also includes other components around risk adjustment, exclusions, some of the things listed here. So, those can also be weighed in your vote.

Testing is one component of validity, so, again, just other things to consider that are related to threats of validity and the overall validity of the measure.

So, take a minute or two for you guys to vote, if you want to think about it, and we'll be just right at break time.

CO-CHAIR ASPLIN: When you are done voting, let's be back in 15 minutes, and we'll start again. Thank you.

(Whereupon, the above-entitled matter went off the record at 11:10 a.m. and resumed at 11:30 p.m.)

CO-CHAIR ASPLIN: All right, in an effort to stay on schedule, I'd ask that we come

on back and get ready to finish up 1560.

Very good. Has anyone else joined us on the phone?

All right, let's move ahead and let's just stick with the order that's working. Joe, if you wouldn't mind adding any comments from your perspective on feasibility of Measure 1560.

MEMBER STEPHANSKY: I think we already covered a little bit about the issue of getting the data from the plans and the methodology for collecting the data, which at least makes it more feasible than it used to be, I suspect, from the plan level, and perhaps more accurate.

The feasibility side, I think there still are going to be some issues with differences in cost among plans, for even getting the XML kind of a set up in place, but it shouldn't be extreme. I think we're actually reducing the costs overall by that approach.

So, on the feasibility side, I don't have anything else to add.

CO-CHAIR ASPLIN: Thank you. Andrea, any comments from the TEP perspective?

MEMBER GELZER: Yes, and there was no significant discussion.

CO-CHAIR ASPLIN: All right, any other comments from the Committee?

Hearing none, let's go ahead and vote on feasibility. Are there any other comments, Ashlie, before we move to that section of the SurveyMonkey?

MS. WILBON: Sure. So, there is -- feasibility is just one vote, considering three different sub-criteria or sub-elements, whether or not the data is readily available, can be captured without undue burden and can be -- there -- can be easily implemented.

So, considering those things, why don't you go ahead and submit your -- or jot down your vote?

CO-CHAIR ASPLIN: Thank you.

Let's move on to our last section on -- prior to overall, on this measure, and Joe, comments on usability?

MEMBER STEPHANSKY: Well, I think that's been kind of the big question by the -- questions that have been raised by the Committee members, is do we have something here that is actionable, or something that may be actionable in the future, depending upon some development of some better outcome measures on the asthma side?

This is more of a toss-up, I think for me. I do have, just as a side comment, that we may want to discuss more in a general way, is the fact that some of the measures that the plan may have, let's just look at readmission measures, Ariel and I were just discussing.

Hospitals, for example, are used to looking at the data from the CMS hospital-wide readmission measure, which has a planned versus unplanned algorithm built into it now.

I think NCQA is looking at initially

bringing in an unplanned versus planned algorithm in 2015. Well, that means that the plans have got to think about well, if I'm looking at the NCQA measure for readmissions and there is something that I need to do here, and I need to communicate it with my provider networks in different ways, and I don't have the same kind of a line-up of those measures, there is potential usability issues there, I think, not with this particular asthma measure, but overall, that we need to consider where we're going to go with that.

MR. HAMLIN: So, I mean, to use your example, it's not RRU, but you know, in our readmissions measure, we align very closely with the CMS/Yale measure for hospitals. Where they don't align is because that is a hospital based measure and ours is a plan based measure, and so, where those differences were necessary, in the RRU space, there really is nothing like this out there.

So, we want to align certainly with

what plans are doing and we interact with plans on a regular basis, to look for improvements, and we continually ask for information from them, from our stakeholder panels, you know, what are you using this information for, and how are you using it, and we try and then turn that around in the community and provide them opportunities for improvement on RRU or you know, we call them drill-down seminars, where how you can drill down in your data and look at --

But to be honest with you, again, I get aggregate data and I know most of the plans have much more sophisticated analysis that they're running on these than I could possibly ever hope to accomplish with the data that I have in-house.

You know, they've pretty much gotten to the point where they don't need my advice anymore on what to do with this information. They've gone well beyond what I could provide them.

So, we just continue to try and provide them what they need that's relevant, as far as benchmarks and you know, and service categories to compare peers and help them provide information and public reporting status and compare it with the HEDIS measures for their stakeholders, but you know, that's really as far as -- as deeply as we can provide them that information.

MR. REHM: Just because I know there is transcripts and all of this, I wanted to -- thanks, Joe.

The readmission measure, by the way, does have a planned readmissions removed from the measure and also has readmissions as index admissions, just like the Yale/CMS model.

So, that's done and -- well, you're correct, that is reflected in the HEDIS 2015 specifications.

To the point of, you know, this issue of -- just like we have in our clinical measures, we are very clear to say this is not

a guideline, you know. The same -- in some ways, the measures -- you know, measures are what they are and at times, I think when you're in a measurement group and that's all you're talking about, you tend to think it's the whole world, and it's the whole environment.

Larry and I were just -- and Ben, talking about the role of communities and leadership and just getting your arms wrapped around a single goal and going after it.

I think a measure like all of our RRU measures, as well as readmissions, because in some ways, they have a similar sort of angle of attack and risk-adjustment, but they -- the reality is, is that you know, plans have their own metrics internal to their own organizations, that are going after many things that they have prioritized as important.

There may be some consistency across the plan environment around those. There may be really focused in some plans, that are not in others. Some people may have black

boxes better than other black boxes.

I do think that by having a standardized approach, audit-able measure that sets the table, if you will, it encourages those plans that -- and I can't think of a plan, why they would not be going after this linkage between cost and quality. It would not seem in their best interest, to not do that, and I think that's why I think so many are doing it.

But some of it is underneath the radar and that's fine. I don't think we have to know that. I don't think we have to -- you know, we are not the entirety or the -- we're not the total community here. We're simply a measure in play, and I'm trying to make it a little bit more pedestrian than it is, but --

MEMBER STEPHANSKY: To me, on the usability side, I like this, in this sense. It's like an old fashion quality improvement measure from 20 years ago.

It points to something that might be a problem, but doesn't necessarily mean there

is a problem, and I think there is still room for those kinds of quality measures in the world, even though we have gotten to the point where we seem to want to be able to put a very precise metric on everything.

CO-CHAIR ASPLIN: Nancy, any comments on usability? Nothing? Okay.

MEMBER GELZER: No additional.

CO-CHAIR ASPLIN: Great. Mary Ann?

MEMBER CLARK: Yes, I'm just looking at the different sub-components of this usability and one of them that keeps sticking out in my mind is improvement, progress demonstrated, if new, credible, rationale.

So, again, I mean, I agree at a high level, this is a useful measure and it's something that the plans can look at from a high level, and then like you were saying, use their own data to drill down into where they can make improvements.

But I mean, isn't there -- as part

of this requirement for usability, isn't there supposed to be demonstrated improvements? I mean, I don't know that we've seen any information. It's all anecdotal, right? I mean, unless I'm missing something.

MR. REHM: So, and this may be a question to NQF and Helen, I may not recall and maybe Ben can help me.

This is -- is the -- the usability requirements for the cost measures any different than the usability requirements in others, so that there is no differentiation.

I think because we created this -- you know, we had to make choices on lots of things, and Ben has been able to articulate the lots of different choices we made, whether it was a risk-adjustment model or whether it was a trendable measure or an un-trendable measure, using real data that comes back to us, so that we can compare plans to each other.

Medicaid to Medicaid, Medicaid Southeast to Medicaid Southeast, you know,

whatever -- however you wish to chop the data, and once we made that decision, we lost trend-ability. It's not trend-able, and when we put the stuff up on our website, we have a big cautionary note, "Please don't try to click on five different years of data, and then draw a trend line and say that's informative," because it's not. It's apples to, I don't know, bananas.

So, I think in some ways, we're -- maybe, if you will, an unwilling hostage to the -- what the criteria here, which is very, very important, very helpful, when we think about traditional, clinical measures.

The nature of the data collection, the nature of what we're trying to show here may not fit that terribly well.

So, I don't think it's a weakness of the measure that was a decided choice we made, and it could be we made the wrong choice. I don't know. I think you could lose as much going into trend-able measure that doesn't use

the kind of data that we're using to validate it, if you will.

So, those are choices that any measure developer has to do, and not all measures can do everything we want them to do, and so, I can understand the frustration here.

MEMBER CLARK: I'll just follow up with a question.

I noticed there was one report though on your website that kind of gave a summary, more of a layperson's kind of summary of how these measures are used and things like that, and I was just wondering, do you -- I mean, do you survey your membership on what care process redesign changes, or anything that they've put in place, to help address any issues they found, you know, in their reports?

MR. HAMLIN: So, the public interpretation of our RRU we put on their website is driven primarily by the other audience, like I said, is very interested in these measures, which is the purchasers and the

consumers, increasingly.

You know, consumer's union uses some of this data for, you know, reporting health plan rankings nowadays, because people are -- people want to know.

That is not driven by, like I said, the other -- the sort of other information that we gather from plans about, you know, how they're using the data and where they're finding quality improvements.

Again, we did a lot of that in the early days, but to sort of help provide best practices for other plans to be informed about how you do, you know, quality opportunities across opportunity calculations using this data, for the most part now, the discussions that we had with plans are around their specific annual results.

So, when they see their snap-shot, and they might be surprised by certain categories, we will work with them individually, to sort of help them understand,

you know. We first make sure it's not a data issue, which it could be, so, we'd continually look at that.

But we try and work with them, you know, one-on-one, and these conversations are all over the map. So, really, there is no model for really, them saying, "Well, this is what we're doing," and I think part of that may be proprietary for them.

They want to use their best practices for managing their costs for their members, and their own really innovative strategies, and not share, because they might lose some competitive advantage. I don't know because they don't tell me that, but that's what I suspect perhaps.

We do continually seek feedback on, you know, how they're using information and some of that does go into those consumer briefings that are posted on their web, again, to help people understand, because they're very complex measures and this issue is not straight

forward.

So, those are -- you know, it's -- we do try and provide as many different perspectives and explanations as possible.

We tried to initially develop an elevator speech of 30 seconds or less to describe RRU and we could not do it. We spent a year trying, and it just would not work.

CO-CHAIR ASPLIN: And along similar lines, to Mary Ann's question, and the comment that Ariel had earlier.

Do you have a sense of what percent of the plans that are reporting on this measure are, taking the next step, using architecture of the measure, running it through an internal attribution model and then going to the various components and delivery system, that are delivering care to their members and representing the data to them in ways that they understand how they're doing, relative to other delivery systems that have members with that plan?

MR. HAMLIN: So, my biased opinion is that they all are doing it. I suspect that, you know, given our experience on the quality measure side of, you know, how the -- like, again, as we have -- you know, how much plans take this information seriously, I would suspect that they're using their -- if they're taking the time and the resources and investing in reporting the RRU, they are going to turn that around and use that information, because they would be kind of silly not to, because it's significant to report.

You know, I just don't see anyone -- it's a huge burden to report these measures. I will be very honest with you, and in order -- you know, to do that, just for the purposes of reporting to NCQA, to me, seems nonsensical.

So, I suspect, and you know, there is a bias because, you know, I think everyone is using my measures all the time, that they are using this information in various different ways.

CO-CHAIR ASPLIN: That has been my experience, and that's why I rate the usability very high, because the plans that you've interacted with as a delivery system have taken the extra step and that information is very valuable, as a delivery system understand for similar kind of conditions in a market, and in that context, a plan can also do a price index that's relevant, without revealing specific data.

So, that is why I think usability is quite high. Bob?

MR. REHM: You know, back to Ben's comment. We always think that everyone is using our measure.

You know, I am sure that there are some folks out there that are. Let's be honest, folks prioritize what they're going to go after, and the things they can go after at a corporate level are highly market-sensitive, highly temporal.

I mean, it's -- you know, they make

a big play and you know, if there is a new Federal program in play, and it's all stars and there is millions on the table, they may just go after that, then go after something else.

So, I never want to have -- assume that everyone is following all the stars that we create for them to follow, because they simply can't. They have to pick and choose, and then they have their own internal priorities, as well, which we respect.

You know, I was thinking when we developed this measure and we added the imaging component, the lab and imaging component that didn't use to exist in it before, in many ways, that was being very sensitive to market appreciation of how to control costs in some of those areas.

So, do I think that health plans without RRU being on the table would look at lab and imaging? I think they would look at lab and imaging. The fact that it's consolidated around condition states and populations can be

instructive and helpful to them.

Again, a measure does not live in its little bubble. The measure -- especially a measure like this, and if we're talking about a total cost of care measure, I can't imagine the discussion, because a total cost of care measure is a total cost, and you know, where is the horizontal and the axis, where do they meet?

So, just some considerations about again, this is part of a larger issue and the fact that areas of interest overlap is a good thing.

CO-CHAIR ASPLIN: Andrea and then Gene.

MEMBER GELZER: I just agree with all these comments, and they -- from a plan perspective, these are usable. There is a significant burden to the plan and significant resource output to deliver the information.

But that said, I see these are a reliable way going forward, to really start talking about network composition in a rational

way, and preferred providers.

I mean, this is -- I'm on the evolutionary scale, where this is a positive step.

MR. HAMLIN: Yes, I think I forgot to mention that of the thousands of -- of the thousand or so plans that are reporting RRU, it's all voluntary reporting, so, there is no requirement that they report these measures.

But there is one-thousand plans voluntarily reporting this information to us, so --

CO-CHAIR ASPLIN: Gene?

MEMBER NELSON: I'm always interested in measured improvement, and I was -- wanted to ask a question.

In golf, if you're in the senior tour, you can see your standing at the end of the year and then you can see your standing at the prior year and the prior year, against the field.

On page 32, there is this graph by

product line, and couldn't this be used year on year, to plot my point in the distribution, to see if I am improving relative to my flock year on year?

MR. HAMLIN: So, what we refer to as scatter-plots do that. We present this information. Again, box plots, people love them and hate them. Consumers don't understand them.

So, we developed that scatter-plot idea with the value equation, which has both equality and the RRU dimension, which effectively presents very similar information at a more consumer level.

We feel that people really -- they've been -- we've tested that and people actually really prefer that methodology.

For me as a scientist, it's not specific enough. It doesn't give me enough information, but you know, again when we're talking about additional enhancements to our reporting, public reporting availability for

the data, right now, we have a very high level, which is the scatter-plots, and an extremely detailed level, which is large CSD files of, you know, extensive information for the plan, and not a lot in between for those that might be curious.

So, we're kind of trying to work, to kind of bridge that gap, about how we can have more sophisticated interactive tools to help. You know, if you want this information, click on a button, get it, without having to create it from the data yourself.

MEMBER NELSON: I followed what you said, but I was referring to a plan looking at itself over time, by product line, and I think an earlier remark said we can't do that, but if you look at the data in a different way, maybe you can.

MR. HAMLIN: Well, we can't because again, the data is calculated based on all submissions for that year, and the prices that are adjusted every year, understand it as

pricing tables.

So, we could have several of those thing constant and present this information, but then it's questionable about the -- you know, how useful that information is because we've held -- sort of artificially held so many things constant, to present that information, and right now, we haven't heard from anybody that they're interested in that because they basically, you know, use our individual benchmarks for their plan at the different levels, and then, you know, dive into their own data.

It's that comparison, that sort of snap-shot that they can see, and then that's where they sort of make their decisions about how much input they want to put into their data.

You know, I mean, we could certainly hold these things constant and present that information. I think it just adds one more dimension that hasn't really been found to be wanted, I think by the field at this point in

time, but we hear that it should be something that should be done, we can certainly look -- you know, dive further into that information.

CO-CHAIR ASPLIN: Okay, with that, unless you have another comment, Nancy. Your card has been up. Okay, would you mind flipping that? Thanks. That's okay.

Let's go ahead with a vote on usability, and Ashlie, if you have any comments to introduce this.

MS. WILBON: Usability, again, is one vote overall for usability, keeping four sub-criteria in mind, whether or not there is a current or planned use of the measure in public reporting or accountability application that the measure -- or the developer has demonstrated that there has been an improvement of cost and resource use and performance over time.

I think that question came up and Bob provided a response to that.

That the benefits of the measure

outweigh the unintended consequences, or potential unintended consequences of the measure, and the final one being whether or not the measure can be de-constructed for transparency for other stakeholders, including those being measured.

So, considering those four sub-criteria, you can go ahead and vote overall on usability and use.

CO-CHAIR ASPLIN: And then we have the overall recommendation for endorsed, yes/no vote. I don't know if you have any other comments, prior to that vote, Ashlie, otherwise I think members can go ahead and submit their vote.

Any other Committee comments? Otherwise, we will be leaving this measure for the time being, and I'll hand it over to Lisa.

CO-CHAIR LATTS: All right, well, I think that's probably the first time in NQF history that a first measure has ever been completed ahead of schedule. Usually, the

first measure is the one that we fall down on.

So, we're now going to have a bit of deja vu all over again, and we are going to go to relative resource use for COPD.

Now, we will have our lunch break at the regular time of MEMBER NAESSENS: 30, which is when the food is coming.

So, why don't we see how far we can get through COPD prior to that lunch break?

We're going to take public comments at MEMBER NAESSENS: 25 p.m. So, we'll hold public comments until then.

So, Matt and Mary Ann are the reviewers on this. Are either one of you going to take the lead? Matt is going to?

All right, you want to start with the first section, any comments that you guys need to make, before we -- okay.

MEMBER McHUGH: So, a lot of this is similar, obviously, to our last measure. So, I'll just try to highlight some of the differences and places where there was

substantial agreement.

Around importance, I think the overall tenor was that this was an important measure. COPD is high cost and very common.

The question about being able to track plans over time was raised, as well, and one question about the intent, and it may go in our scientific discussion, was about mis-diagnosis and the ability to truly identify patients with COPD and false-positives.

CO-CHAIR LATTS: All right, any other comments? Mary Ann, anything you want to ask -- add to that?

MEMBER CLARK: No, I think that about covers it.

CO-CHAIR LATTS: Andrea, any comments?

MEMBER GELZER: No additional comments.

CO-CHAIR LATTS: All right, any other comments from the Committee, prior to voting on importance to measure?

Seeing none, everybody, if you could go to the SurveyMonkey, enter in COPD measure and vote on importance to measure.

MS. WILBON: I'll just highlight again that importance to measure includes three sub-criteria, that we have you vote individually on each of those three sub-criteria, which is a little bit different than the others, and then overall vote for importance.

So, keeping in mind, the kind of three different components around opportunity for improvement, the measure intent and whether or not it's a high priority area with demonstrated variation in care.

CO-CHAIR LATTS: All right, if everybody is finished voting, we will move on to scientific acceptability. Matt, you want to take that?

MEMBER McHUGH: So, on -- and we're doing reliability first?

CO-CHAIR LATTS: Correct.

MEMBER McHUGH: So, there were some questions raised about whether there actually was information presented at both the data element and performance measure less. I don't know if particular people who have questions about that, want to follow up.

There was a question about patients with both COPD and asthma, who were in -- are included only in this COPD cohort.

The question about the Medicare H -- I'm sorry, I'm going into validity. I think those were the biggies.

CO-CHAIR LATTS: Mary Ann?

MEMBER CLARK: Just one other thing I wanted to point out.

It looked as though for identifying the COPD patients, that they -- it was any time during the measurement year, correct, as opposed to the asthma one, which was two -- or prior to that.

So, I guess what I'm wondering is, how does that affect costs if you only -- you

know, if the patient was identified, you know, just at a certain point in time in the year, they were a new patient, I guess. Is that --

MR. HAMLIN: So, couple things. You're correct, that COPD and asthma concomitant and diagnosis patients are included in this data set, where in asthma, we exclude specifically COPD to try and keep it a little distinct, if you will.

You know, because COPD also includes chronic bronchitis and other sort of diseases that I think were clinically deemed to be appropriate by our measures.

We did look at many years ago, on the HEDIS side of the clinical COPD measures, looking at the diagnosis criteria for COPD. We did extensive inter-rate reliability testing on the -- you know, the piece of information in the medical record, translated down to the ICD-9 level codes for COPD and chronic bronchitis, etcetera, etcetera.

Those were found to be in very high

agreement, and to your point, again, I think that, you know, because these again, are plan measures, the number of people who are included in the measure, over the measurement period is fairly consistent because of the fact that these are very extensively risk-adjusted through 13 different risk categories and so, patients starting late in the measurement period would only have so many services included for that, you know, once they're identified.

But we would assume that difference is not great enough to affect the measure between plans, if you will, because of the way that the measures are structured.

So, one plan is fairly comparable to its plan -- other plans because of -- you know, because of the adjustments for -- as much as we can adjust for it, to be honest with you.

CO-CHAIR LATTS: Andrea, any TEP comments?

MEMBER GELZER: Yes, that

discussion also occurred at the TEP level, and NCQA confirmed that the diagnosis code, if the diagnosis code is present in the record during the measurement year, they're included in the measure population, and that seemed to then be okay with the TEP.

There was some discussion of the potential of mis-diagnosis of COPD at the technical committee level, and they talked about individuals with sleep apnea, but again, that was just a minor discussion and didn't seem to be a real issue.

MR. HAMLIN: Actually --

MEMBER GELZER: And I believe that NCQA said there was good correlation.

MR. HAMLIN: I just wanted to actually add one more thing to my response to Mary Ann.

The measures are reported on member-month basis too, so member-months to member-months I think is much more consistent than inclusion in all costs. So, sorry for

that.

CO-CHAIR LATTS: All right, any discussion on the reliability, and I heard some phone -- is there anybody on the phone that wants to make a comment? It seemed like there was some action there.

MEMBER DAMBERG: Yes, this is Cheryl. I wanted to ask NCQA, because I wasn't clear from the documentation.

You indicate that you only report on plans with at least 200 cases, and I know you stratified the results. So, is that 200 cases per strata?

MR. HAMLIN: I'm sorry, I missed part of your question. Can you repeat it?

MEMBER DAMBERG: Sure. So, in the documentation, it seemed that you're only reporting out for plans with at least 200 cases, and I wasn't sure whether that 200 cases in each of the strata that you report out on.

MR. HAMLIN: No, the population has to be at least 250 members, 200 members for the

plan to be able to report, or for us to report the data for the plan.

So, it's not episodes. It's not cases. It's just members who have been identified with that disease. It's not by each of the strata, because that would be a huge population.

CO-CHAIR LATTS: All right, other comments? Yes, Ariel.

MEMBER BAYEWITZ: I am not a statistician. I was just looking at page eight here with the results, in terms of percent of plans with no more than one quartile shift, just getting to reliability.

Just seemed like a high percentage of plans shifting from, you know, in terms of moving to quartiles, you know, 30, more than 30 percent in one of the buckets, commercial PPO. The shifting two quartiles.

So, I mean, how do you -- it just seems very high. Is that not statistically a problem?

MR. HAMLIN: So, there are in some product lines for this measure, in the earlier years, I think, if you make sure you're looking at the final column, which I believe is the most recent data.

The quartile shifting in those decreases significantly, but it is -- there is a proportionate plan, you know, 15 to 20 percent in some cases that do shift, and again, this is an issue of, you know, we look at those plans, we look at the outliers, if you will. They're not really outliers if they shift more than two quartiles.

But it is sort of the rough trending analysis that we can do, to see if there are significant changes between plan performance from year to year, and some of that is driven by plan changes. Some of that is driven by the number of plans and the types of plans they submit.

It again, was -- you know, again, since the majority of plans -- it really is

dependent again, and there were plans in their first completed reporting, that may shift more because of, you know, just coming into the RRU reporting methodology, if you will. Sometimes, you know, the earlier years, our data is used and sometimes there are actual -- more representative of the results of the plan.

But I think to your question, you know, we do expect these data to shift year to year to year, and so, I don't think that, you know, a plan shifting more than two quartile -- and we say more than two quartiles because much of the shift is around the meaning, depending upon the -- you know, so they could shift one quartile easily. They can shift between the mean, just depending upon the submission.

So, we don't treat them as outliers, but we can just look at those plans carefully, to try and understand why they're shifting, but those numbers have gone significantly down in the years, and now, we're seeing, you know, again, for most plans or for most product lines,

it's really in the high 80's and low 90 percent of plans who are not shifting more than one quartile.

So, it's really the only -- it's the only trending analysis we can do and look at to see, you know, changes in plans, but so it's a little rough in the methodology, but it's -- like I said, because we only get aggregate data, that's all we can do.

CO-CHAIR LATTS: Other questions or comments? Mary Ann?

MEMBER CLARK: I think there was just another comment that someone had, regarding -- and maybe you can clarify this, but about whether denied claims were included in identifying the population. I don't think they were, is that correct?

MR. HAMLIN: Denied claims are included in the population. You know, most of the -- the costs, the PMPM, you know, resources used are used -- are the only ones used for the resources that are tracked.

So, you know, for identification of population, denied claims are included, but for the resource use reporting, only claims that are expected to be paid or paid are included as part of that, because the plan did not actually pay the claim.

MEMBER CLARK: Okay, then maybe, I don't know if the person who had that comment wants to elaborate or question that further, because I think the rest of the comment had to do with maybe the service wasn't really -- maybe -- if it was a -- as a result of a test, for example. Maybe it was, you know, a rule-out, COPD, but I don't know. That seemed like that was the area of the question.

MR. HAMLIN: I mean, again, you know, we include the denied claims in the ID algorithm, because we want to make sure that we're capturing as many, you know, potential COPD patients as possible, to try to be as sensitive as possible.

But again, you know, it's resources

used to manage a chronic disease population. If those claims were denied, I think our opinion is that if they were denied, they were probably denied for a reason, because they weren't necessary for managing that condition, and on a plan by plan basis, you know, that's kind of the assumption we have to operate on, basic level of information we have.

CO-CHAIR LATTS: Brent?

CO-CHAIR ASPLIN: I think that Fred, if I recall correctly from the TEP conversation was, is that if it's used to get people in to the measure, then there are no resources used -- associated with the plans, with the disproportion high rate of denial, going to look better from a relative resource perspective. I don't know if you can comment on that.

MR. HAMLIN: I think in theory, they probably could, but given the fact that there are 13 risk categories for multiple age and gender cohorts for this, you know, we would

see -- the HCC would assign plans to -- you know, those patients would end up in a -- probably a lower risk category first of all, which would make us suspicious in the first place, and then they would also sort of end up in the zero utilization and we look at the zero numbers very, very carefully.

So, I think in theory, it's possible, but I think, you know, does it -- a vast proportion, it could affect the measure, probably not.

CO-CHAIR LATTS: Gene?

MEMBER NELSON: I think in the prior measure I saw the information on extreme outliers greater than three and less than .33.

What is the policy for outliers here and how many plans hit the outlier above or below?

MR. HAMLIN: It depends. There is a table for this measure too. I have to find it for you, but I can explain.

Outliers are -- is a very low

proportion nowadays. Initially, it was -- and primarily we find the outliers are in calculation errors. They're not actually in relative resource use. They're not in outliers of relative resource use. It's generally in programming and calculation errors and reporting the measure.

New plans are the ones most likely to end up in the outlier status, but less than one percent for almost all of measures nowadays are plans that are identified as outliers.

Early days, that was not the case. It was much higher, and so, we use that outlier status check to sort of look to see who we should look at carefully, to see if it's -- if they're having problems reporting the measure, which is usually the case.

MEMBER NELSON: Is the same convention used, .33 and --

MR. HAMLIN: Yes.

MEMBER NELSON: -- bigger than three?

MR. HAMLIN: Right, and those outliers are not included in any of the calculations because we think it's a data issue. We don't think it's an actual resource reflection, reflection of resources use.

CO-CHAIR LATTS: Any other comments on reliability?

If not, let's move onto validity, and just plow through.

MEMBER McHUGH: Again, some of the comments were about the presentation of the materials, and being able to find all of the necessary pieces around particularly in the empirical testing of the measure. We did say this earlier, but there was one comment about the R-squared of the model that was being used, not being evidently clear.

This may be less of an issue with this measure than the prior one, but again, the comment about the HCC, using the Medicare HCC's and I believe this population is 40, and what -- the degree to which that's applicable.

Again, as with all of our measures, this is -- uses billed services and denied claims as well, but and whether that's actually measuring the process of care that is being provided, but is probably the best that we can do.

I think those are the primary areas, different from the prior measure. Mary Ann?

MEMBER CLARK: I mean, there were some comments, I think on the asthma one, as well, but on this one, on severity, you know, COPD severity which, you know, if we're using claims data, the best we can do is adjustments with the diagnosis codes, primarily.

So, until that changes, I don't know that there is a lot that can be done to -- you can only use those codes to proxy, sort of as a proxy for severity. So, I don't think there is a lot that can be done.

MEMBER McHUGH: And it's probably just for the record, to note that comments around SES adjustment were raised with this

measure, as well.

CO-CHAIR LATTS: Right, we are still are bound by the previous -- the rule -- the law of the land today is do not adjust, whether it's that law of the land tomorrow, we'll see.

Andrea, any TEP comment?

MEMBER GELZER: No really significant ones.

CO-CHAIR LATTS: All right. Any Committee comments on validity? Everybody is quiet.

MR. HAMLIN: Actually, I just wanted to address the R-squared comment a little bit, because I don't think I fully addressed it enough.

You know, again, we use R-squared to test the appropriateness of the HCC model for the RRU measures, you know, but again, since we don't get member level data in every year, it's very, very difficult for us to retest that model every time, and until additional guidance or,

you know, some other appropriateness criteria steers us in another direction, we probably won't retest the HCC appropriateness for this model because it's been tested and found to be, you know, most relevant and valid.

So, you know, again, even though it's a huge amount of data coming into us, it's still at the member level. So, it would be very difficult for us to get do -- sort of annually test the R-squared.

So, we annually test as much as we can, to make sure that relevance of the model is, you know, still working and we listen to the public feedback continuously about that, but without big, big, big data --

CO-CHAIR LATTS: All right, any other comments? Anybody on the phone have any comments?

Have Jennifer or Martin joined us, by any chance?

All right, if not, then let's go ahead and vote on scientific acceptability,

both reliability and validity.

All right, well, we are going to plow through and finish this before lunch.

So, let's go ahead then and move onto feasibility.

MEMBER McHUGH: Really seemed, at least from the comments, pretty much agreement on feasibility.

MEMBER GELZER: I mean, similar, it was -- you know, there was a lot of discussion that this measure was constructed similarly to the asthma measures, and quite honestly, I think the TEP Committee had a little bit more issue with the asthma measure than the COPD measure.

So, there were no other significant comments.

CO-CHAIR LATTS: All right, any comments from Committee? If not, let's vote on feasibility and usability.

MEMBER McHUGH: So, here, again, this is a measure that is in use. There was

discussion about how plans would and would use it, but again, in line with our prior discussion. Yes, I think --

CO-CHAIR LATTS: Mary Ann? Andrea?

MEMBER CLARK: No other comments. They're pretty much the same, I think, as for the asthma one.

MEMBER GELZER: Agree.

CO-CHAIR LATTS: All right. Somebody has something exciting to say. Nancy?

MEMBER GARRETT: I don't know if it's exciting, but I did want to bring up the issue of stratification for SES. I don't know if this is the right place or not.

But you know, when we did that vote on the Phase 2 measures, we added a recommendation that the measure be stratified for SES variables, and we could do that here for this measure and the previous one, I think.

So, do we want to do that, and

realizing that data is a big problem, but with the recommendation in place, then it can start to move that ball forward, in terms of trying to find the data and finding ways to do it, so that's kind of an open question for the Committee.

I would recommend it, based on the evidence that we've seen, that with the asthma and the COPD, it would be very important to include.

CO-CHAIR LATTS: And you have the inside track on what the recommendations are going to be, so, we're anticipating that the recommendation --

MEMBER GARRETT: They're getting ready.

CO-CHAIR LATTS: -- will be to stratify.

MEMBER GARRETT: Yes.

CO-CHAIR LATTS: Brent?

CO-CHAIR ASPLIN: Yes, I am supportive of recommending it. I think whether

you put it in the model, the risk model or if you stratify it, you've got to have the data either way.

So, I mean, I think the same data limitations are going to hang up, making that something that can be implemented over the short-run, but that is by no means, a suggestion that we should not recommend it, you know.

CO-CHAIR LATTS: Larry?

MEMBER BECKER: Yes, I am sort in the same place, and you know, in light of these being relatively accepted, non-controversial measures, as we start to think about the stratification and the SES, might be a good place to start and to learn, and sort of shake out some of the issues, so that we can learn from that and it's something that's less controversial, just because it would be a good opportunity, because we will get to things, I'm sure that are more controversial.

CO-CHAIR LATTS: Matt?

MEMBER McHUGH: Just a point of

clarification in terms of what we might recommend and what we recommended previously.

I thought that we recommended to report both stratified and the full process spectrum, un-stratified, essentially. Yes, yes.

MEMBER GARRETT: The current NQF guidance is that the endorsed measure can't be statistically risk-adjusted.

MEMBER McHUGH: Right.

MEMBER GARRETT: And so, we recommended that since we can't recommend that, that the measure be stratified after the fact, meaning reported by the relevant socio-demographic groups, to be able to identify disparities in care and adjust them.

So, that would be more, as the measure is used in the real world, recommendation for after, it kind of leaves us.

MEMBER McHUGH: So, that users would only see their -- within their strata, the results within their strata or would they see

all of the data?

I mean, that's my -- I would just be interested in making sure that the full spectrum, as well as within the strata, are available.

MEMBER GARRETT: So, the NQF endorsed measure would be unadjusted for those variables. So, that would be the one in use.

MEMBER McHUGH: Okay.

MEMBER GARRETT: But then we'd be making recommendation that it also be stratified when it was being used out in the real world by the relevant system --

CO-CHAIR LATTS: In a non-NQF-endorsed way, as of today.

MEMBER GARRETT: Well, no, the stratification can be endorsed by NQF. It's maybe not the word endorsed, but the current guidance actually recommends that NQF send measures out into the world with this stratification idea, as a way of dealing with the SES issue, and yet, it hasn't happened very

often, because I think committees have just said, "Well, we can't risk-adjust, so we're not going to worry about it," and also then, once the measures leave us, you know, there is not a lot of control that NQF has over how they're used.

But we can make a strong statement, I think by putting that in the recommendation in a strong way, to help people think about it more. So --

CO-CHAIR LATTS: Bob, do you want to comment on that?

MR. REHM: Sure. So, a recommendation, you know, to essentially alter the implementation of a measure is not an insignificant task.

I mean, you're asking, and I think we've reflected throughout and I think Ben has in particular, that we're operating kind of in a learning environment, where we're open to the suggestions of this Committee and others, to improve all of our measures.

So, you know, we'll obviously -- this topic is getting a high level play, across the healthcare field, and within all of our own committees, and we're doing it in a thoughtful way.

In many ways, you're suggesting that we test the hypothesis, that for a measure like RRU, that you know, we already offer three stratifications. We offer Medicaid, commercial and Medicare, and let's be honest, that's not -- that's a lot more than many, and it is instrumental and it does affect policy.

So, I think from a -- to say that we're -- and I know no one has said this, but to -- we're not operating with dis-interest. We're operating with a high level of interest.

You know, Ben pointed out that in this particular measure set, and I appreciate Larry, that you suggested this is kind of a nice place to start, because it's non-controversial and easy. It's 70 -- how many data elements? Seventy-thousand?

MR. HAMLIN: Between 69,000 and 75,000.

MR. REHM: Seventy-five-thousand data elements. We literally had to invest thousands of dollars in our servers to be able to accommodate the input on this measure alone. It was a revolution at NCQA. We didn't know what hit us.

We felt like healthcare.gov, but the early version. So, I just want to make sure that you appreciate that we are very sensitive to this. We appreciate recommendations.

I do think that there is somewhere, a fuzzy line between letting us know what's important to you and then, there is that fuzzy line, the gray zone of saying, "This measure is a better measure, if you do x, y or z."

Sometimes, we know that, and we kind of have an instinct about that. Sometimes we have an instinct that that's probably really a bad idea, you know. I am not going to characterize our instincts on this one.

But I think that it's a hypothesis that we may want to test, you know, I guess it's the difference between a -- being part of a larger policy conversation and using the measure as a lever, and I think that I can appreciate why you'd want to do that. It's the famous nail-hammer thing, you know, everything is a nail.

But I just -- I am suggesting and saying, you know, we're reasonably conservative about our measure development process. It's taken us years to get to where we're at with this measure and with other measures.

So, I think that we would want to test those things, because I don't think this is a one-size-fits-all answer, but we've heard certainly, what you said, and I just thought it would be helpful to get our reflection on that larger issue, and to separate the measure, which is trying to live and breathe its own space, and the larger, you know, political,

legal policy sort of orientations that we all have to improve the healthcare system and make it more sensitive to patients.

CO-CHAIR LATTS: Thanks, Bob. Those are measured comments. Appreciate it. Cheryl on the phone wants to make a comment.

MEMBER DAMBERG: Yes, I just wanted to -- you know, we've looked at the stratification piece, because I do think it's an important step in the right direction, in terms of separated by Medicare versus Medicaid versus commercial. So, I applaud NCQA for doing that.

I do think that, you know, moving forward, it's not going to be in the near-term period, but I do think for plans, as well as the provider networks, that they operate, it's going to be essential to have more drilled down by different race, ethnicity, populations because there is where we see the disparities.

CO-CHAIR LATTS: Mary Ann?

MEMBER CLARK: Just a comment, an

observation.

I agree that with the stratification approach, but I thought we also heard that a lot of this information is not collected or it's very difficult for the plans to collect, and you know, may not be realistic to have it reported out that way, at this point, I guess. I don't know.

MR. HAMLIN: Yes, I mean, to simplify it down, I call this the Costner principle.

I mean, if you build something and build a template, they eventually will come, but of course, all they had to do was bulldoze a corn field. We have to do a lot more, and I agree, I think with the comments that, you know, if we don't recommend this and we take -- you know, we take NQF Committee recommendations very seriously and we try and work them in wherever possible into our process and we, you know, move it forward.

But you know, so the

recommendations that it should be done this way is one thing. You know, the requirement that if you want to report this measure, you have to report it this way is another, and I think we're still trying to find the carrot, not the stick, at this point in time, as far as SES goes, or even, you know, the geo-coding analysis, because I think the plan -- the data is just too unreliable at this point for our comfort level and we are very conservative, before we release a measure. We have to be extremely comfortable with the comparability of the results.

CO-CHAIR LATTS: Nancy?

MEMBER GARRETT: So, two points. One is if we make a recommendation that this measure should be stratified by the socio-demographic characteristics that are relevant, it gives you leverage with the plans targeting the data. So, it could actually help you.

The other thing is, that I think it is more of a recommendation than a requirement,

that I guess I look to NQF to speak to that, because I don't know exactly what that would look like, because it's not really part of the endorsement, it's more part of the -- isn't that true?

DR. BURSTIN: Yes.

CO-CHAIR LATTS: Because I don't think we would have the capability to require it, because it's not the measure that we're reviewing, that we would have to review -- there would have to be testing and the supporting information, that we would need, just like we do with --

MR. HAMLIN: But you're actually right. I mean, it would be, you know, this Committee making the recommendation that helps us continue to drive this importance for it, and you know, and folks like CMS pay very close attention to NQF recommendations.

So, that also helps. So, if you're recommendation, then you'll hear about it and pick it up and it just continues to move that

conversation forward, if nothing else. So, it's definitely helpful.

CO-CHAIR LATTS: I think it feeds into that conversation we had this morning about wanting to be very explicitly clear about what it is that we want as a Committee, and if this is something that we want, making the recommendation, I think is helpful and you know, getting the resources and all that.

So, if there are no further comments, let's go ahead and vote on usability. Which button?

MEMBER GARRETT: Just a process question. Can we vote on that recommendation at some point or do we --

CO-CHAIR LATTS: Do we?

MEMBER GARRETT: How does that get done?

MS. WILBON: So, we can document it in the report. It would be -- so, the measure, depending on what the final recommendation on the measure would be, for example, if the

measure was recommended, we would have notes with the measure -- what the recommendation said that the Committee also recommends this measure be stratified and used, but it's not dependent on the endorsement recommendation. I think it's the clarification that I think Ben and Bob were looking for.

So, if you guys are amenable to that, we can just document that the report, that that was kind of the Committee's feeling and they felt strongly that it should be done, but it would not change your recommendation for endorsement, if that makes sense.

MEMBER GARRETT: I don't know if we've heard from the Committee. I mean, we've only heard a couple comments.

MS. WILBON: Yes.

MEMBER GARRETT: So, I was --

CO-CHAIR LATTS: I was wondering if we should -- yes --

MEMBER GARRETT: Okay.

CO-CHAIR LATTS: -- get some sort

of a, I don't know, straw poll.

MS. WILBON: Sure, we can.

MEMBER GARRETT: My proposal is really for this measure and the previous measure.

CO-CHAIR LATTS: Yes, so maybe, everybody do their online voting for usability, as well as for the measure endorsement overall, and then we can do a hand-raise for the stratification by socio-economic status.

All right, looks like everybody is done with that.

So, do you want to state your motion for the record?

MEMBER GARRETT: So, I move that we include in the -- that we recommend that this measure and the previous measure be stratified for socio-demographic characteristics that have a conceptual link and empirical evidence of a link with both of the variables.

CO-CHAIR LATTS: Do you need a second? Do we have a second? Andrea?

MEMBER GELZER: Can I modify that?

CO-CHAIR LATTS: So, but it would not -- I just want -- it would not delay -- the request will not delay implementation, should not delay implementation of the measure.

MEMBER GELZER: It is unrelated to --

CO-CHAIR LATTS: I mean, it's --

MEMBER GELZER: -- endorsement of the current --

CO-CHAIR LATTS: Correct.

MEMBER GELZER: -- measures before the Committee?

CO-CHAIR LATTS: Thank you for that clarification.

MEMBER GELZER: Then I second.

CO-CHAIR LATTS: Okay, all in favor, please raise your hand and for those on the phone, if -- do you want them to email you?

(Off the record comments)

CO-CHAIR LATTS: Chat. Please chat yes or no. All right, any opposed?

Larry, are you opposed? Okay, he just feels really strongly in the affirmative.

Okay, we're good. All right, okay, so, at this point, let's open it up for public comment.

OPERATOR: At this time, to ask your question, please press star and then the number one.

There are no questions or comments at this time.

CO-CHAIR ASPLIN: All right, excellent. Well, guys, great work on the Committee. We are ahead of schedule and it is time to break for lunch.

MS. WILBON: Yes, so, we'll take 30 minutes. Yes, we may extend, but I think we'll do a 30-minute lunch for now, and so, we'll come back at one o'clock. Thank you.

(Whereupon, the above-entitled matter went off the record at 12:30 p.m. and resumed at 1:05 p.m.)

CO-CHAIR LATTS: So, we're not yet

sure when the developers will be here for the next phase, so in the interim what we thought we would begin with, courtesy of Nancy, is a debrief of the decisions and the concerns around the Phase 2 measure that is similar to the pneumonia measure we'll be considering in a moment so that --- because if you recall, it was very controversial at the meeting itself. We ended up in that --- well, I think initially we even ended up with a less than 40 percent to not continue, and then there was a re-vote at the meeting, and we ended up in the equivocal zone, that 40 to 60 zone. Then we had the follow-up call a couple of weeks ago, and then there was concerns --- and then there was a suggestion to re-vote, and then there was a suggestion not to re-vote, so there was some back and forth about that.

Ultimately, there was a re-vote. Enough people changed their votes, which was probably two to three people that we are now in the --- squeaking by in the endorsed range. So,

Nancy had a suggestion, and I think it's a good one, to talk a little bit, debrief in this Committee. And we're obviously a smaller group than we were at the last meeting, but just see if anybody had any comments about their feeling on the measure, where we ended up, again, in anticipation of review of a very similar measure in a few minutes.

MEMBER GARRETT: I think the moral is just keep voting and then ---

CO-CHAIR LATTS: Words for the federal government to hear.

MEMBER GARRETT: Can I start with -- I think just my general comment or thing I want to talk about is that it feels like in our in-person meeting we had raised significant concerns with the measure, so do people who changed their vote feel that those concerns were addressed in the interim by the developers, or what changed so that now we're endorsing that measure, and then that's going to inform the next conversation? So, that's

what I'd like to hear from people about.

CO-CHAIR LATTS: Anybody want to comment? It certainly may be that the people who changed their vote are not here today.

MEMBER WALKER: I would like to comment because I did change my vote. I would say that the --- in our initial conversation and present conversation I was quite dissatisfied with the risk-adjustment approach. And when we were starting the evaluation for the pneumonia measure clearly it was the same issue that was arising, so I started to do a little bit more digging, reading a little bit more about what they had done for their Quality Measures.

Now, one of their responses had always been --- the developer's responses has always been we're limited by the data so we can't do very much around the severity adjustment. And, also, another response was, you know, we did the same methodology with the Quality Measure and that was approved. And, for me, I found those two responses quite

insufficient. But if you read a lot of the background information that comes with the --- you know, all the references to all the previous testing and validation that they had done around the Quality Measures, they actually had done quite a bit. So, in the Quality Measures they had compared the use of claims-based data to data that was extracted from medical records. And for, I think maybe it was the heart failure or AMI, I can't remember, they were able to report it at the hospital level. For the other measure and for pneumonia they could only report it at the state level because of limitations with the data.

But reading those reports was very compelling to me, so I felt that given what they had done to demonstrate that the use of claims-based measure was very similar to what you would achieve using medical --- data extracted from medical records, I felt that it wasn't as much of a concern now around the risk-adjustment, so I changed my vote.

CO-CHAIR LATTS: Thank you. That's helpful. Other comments? Yes?

DR. BURSTIN: I would also be interested to see whether the public comments also modulated your thinking for those of you who thought differently after the call.

CO-CHAIR LATTS: Nobody is commenting. Lina.

MEMBER WALKER: I would say the public comments didn't really. I mean, they echoed all our concerns. There was nothing really new in the public comments.

I would say that the way the developers respond to some of these questions really don't do them justice. I mean, they can anticipate, I think, some of our concerns prior to coming to these Committees, and in my opinion ---

CO-CHAIR LATTS: Can you get a little closer to the ---

MEMBER WALKER: All right. Is this better?

CO-CHAIR LATTS: Yes.

MEMBER WALKER: Right. They can anticipate what our concerns are. And it makes sense to me if I were a developer to address those concerns head-on. When we write articles we know what the data limitations are. We talk about them in the article and say why we think it doesn't affect our results.

They don't actually do that when they're writing up their reports. They just say this is what we did, instead of saying we did this, we realize this might be an issue, this is why we think it isn't an issue. And then it can alleviate a lot of the concerns, a lot of the discussions around the Committee table.

CO-CHAIR LATTS: I actually thought C-- you know, we'll get to it whenever we get to it this afternoon, but I thought the developers did a much better job this round anticipating our concerns than in the second version.

Joe, did you want to make a comment?

I saw you motioning before Nancy ---

MEMBER STEPHANSKY: I think we talked once before at one of our meetings about the idea of the developers giving us more of a story about how they got to where they are, and how useful that could be. And I think we saw a little more of that in the pneumonia piece than we had earlier. That story I think is very important, and I think that's what you were getting at.

CO-CHAIR LATTS: Well, I agree. It gives you context and it gives you an understanding why they didn't do some of the things that we've asked them to do. Nancy?

MEMBER GARRETT: I think my comment --- I didn't change my vote so I still have concerns about the measure. And I think the concern is still applicable to the next measure. And one of the main ones is really about the risk-adjustment, the conceptual risk-adjustment, so not so much the technical. But they're not adjusting for co-morbidities

that happen in the hospital that could possibly be related to a complication caused by the care. And that conceptual framework means that there's a lot of things they're not risk-adjusting for that actually are patient characteristics and not caused by the care. So, it's like an over-correction.

And I don't have a solution for it because it's one of the --- it's kind of a --- it's related to data limitations, but it's also conceptual, so coming from a provider organization that really concerns me that you're not adequately adjusting for the patient severity and what's really going on with that patient. So, that's a big concern I have with the approach to the risk-adjustment, in general.

MEMBER WALKER: Can I just respond to that? I mean, the issue is whether or not there are systematic differences across hospitals. Right? So, if they're making the same kind of --- doing the same mismeasurement for one

hospital, potentially they could be doing the same mismeasurement for the other hospitals. So, I guess the question I would generally ask myself is, is there anything for me to suspect that they're more likely to be mismeasuring for one hospital rather than another?

I mean, I don't know the answer to that question, but it's the same kind of conundrum I had with the severity issue, which is similar to what you're saying. I think it's a --- so, I mean, it would be helpful for them to respond to that question, whether or not they see systematic differences across these hospitals, which is really what we're trying to get at.

MEMBER GARRETT: In some ways that's kind of unmeasurable, but yes, that is the question. And I think the --- just conceptually that I would imagine that there is a lot of differences in patient severity across hospitals because there's so much difference in patient needs. So, I'm going in with the

hypothesis that there probably is and that we're not capturing it.

CO-CHAIR LATTS: And my issue with that then ends up that you let the --- to my mind, let the perfect be the enemy of the good and that if you wait until we can get all that information because of the limitations of our data, we wouldn't be able to do anything. So, I would rather have something knowing it's imperfect than nothing. Any other comments?

MEMBER NAESSENS: Lisa, just in responding to your comment.

CO-CHAIR LATTS: Yes, Jim.

MEMBER NAESSENS: That would be good if we knew what the measure intent was, and if they've got limitations on it. But when it starts coming into accountability measures, then we have more concerns that the perfect or the good may not be good enough --

CO-CHAIR LATTS: Right.

MEMBER NAESSENS: -- to not bring in some unintended consequences.

CO-CHAIR LATTS: Yes, and I think that gets to an issue that we've talked about before on this Committee, which is the -- how divorced we are from any control over how a measure is used. So, whereas, we might -- we would be more in favor of a measure used for Y than used for X, we can't control that, so it's an all or nothing from that perspective. Yes, Nancy?

MEMBER GARRETT: I was going to quickly say, I mean, this is going to be used in value-based purchasing, so let's just be open. This can be used to move large sums of money around so you absolutely have to think about those consequences.

MEMBER CLARK: Playing the devil's advocate a little bit in maybe the way they're thinking. Of course, I can't speak for CMS, but they're trying to really hold the hospital accountable for these costs even out to 30 days, so from their point of view probably from what occurs when they enter the hospital, if they've

had a complication, you know, as a result, the theory is that's something that occurred while they were in the hospital, so that's maybe the issue behind that. That, you know, also, I'm just thinking from my days in looking at clinical trials. I mean, that would be considered an outcome of an intervention, so that's kind of a standard way to approach these things.

CO-CHAIR LATTS: John?

MEMBER RATLIFF: I would be concerned on those comments that CMS is going to apply these to like the letter grades, they're going to apply a Hospital Compare, so our kind of rarefied decision here is going to become like a grade of a B, or a C for a facility. And let's worry about the implications of how this will be used.

CO-CHAIR LATTS: Carolyn?

MEMBER PARE: And my concern isn't so much about the enemy of the good, it's back to the is it good enough? And is it something, if

there's so little variability in the scores, so little variation in the scores because you've made all these exclusions and changed the specification in such a way what's the purpose? How are you really going to get to true -- being able to discern the actual value between the organizations if they all look alike?

CO-CHAIR LATTS: All right. Any other comments? Let's do a quick check of who's on the phone. Cheryl, are you on the phone now?

Herb, are you on the phone?

MEMBER WONG: Yes, I returned about 10 minutes ago.

CO-CHAIR LATTS: Great. And then do we have Jennifer or Martin? Okay, so just Herb right now.

Okay. So, the developer will be here at 2:00, so in the interim we thought we would get a jump on tomorrow's discussion. And this is the part where we, as a Committee, will give feedback to NQF about what we want and what's going on within the cost and resource use space.

So, if everybody wants to turn to their Discussion Guide handout, are you guys going to start?

MS. WILBON: So, for those of you that have had a chance to review ahead of time, a lot of the front matter, I would say in pages 1-3, is kind of background. It walks you through kind of a timeline of our work. Given that this was intended to be kind of a strategic discussion, we wanted to make sure that everyone, although we talk about it all the time, had a good kind of perspective and idea where we've been. So, the first three pages are really devoted to that.

I want to focus your attention on page 4, though, where we've laid out a table of the various different topic areas that we currently do have topics -- clinical and topical area committees around on our Quality side. And some of the work that we've been doing with this group and with the linking costs and quality work has been grounded in the fact that,

you know, are looking for resource-use measures to be linked with quality measures in some fashion; again, still exploring what that is, but the idea would be, you know, resource-use measures, or quality measures to accompany the resource-use measures that are endorsed.

So, to that end we took this chart to give you a sense of the measures that are endorsed in the other topic areas, and kind of where we are right now with cost and resource-use. Obviously, the numbers look a little disheartening because, obviously, we don't -- the portfolio of cost measures is not that large. So, we've just kind of done a side-by-side here of quality measures, where we are with resource-use measures, just to kind of give you an idea of where we are and get your thoughts around where we need to go. Do we need to have a resource-use measure in every clinical topic area? Is there another approach that we should be using to grow the portfolio?

I also wanted to draw your attention

to Appendix A, which -- let me just look and see what the page number is, where we've captured some of the recommendations that came out of the MAP Affordability Task Force. This Task Force is one that we've talked about in I think our last meeting, but to give you a quick refresher, they were tasked with coming up with coming up with what we have affectionately termed a family of measures. But because there aren't a lot of measures around cost, their exercise tended to be more around a gap analysis, if you will, identifying different topic areas for high-impact areas that impact cost, or spending within the health care system, and identifying those measures that currently exist, and then measures where they think end in concepts or ideas for where they think measures should exist to further the field in terms of cost measurement.

So, laying that out we wanted to, again, get the Committee's input on where we think we need to go. We don't currently have

funding for a next phase of measure endorsement at this point, so we wanted to really take the time to think through if you were to try to further this work what should that next step be? We've had very low numbers of cost measures that have been submitted up to this point and to the discussion that occurred this morning around, you know, not getting the measures that we want, and how we can, you know, strategize about getting that. And, specifically, during this discussion I would encourage you guys if you have ideas, to Helen's point, about specific measures and not just, you know, here in this table in Appendix A you'll see there's kind of some high-level conceptual ideas about conditions like diabetes. We need condition-specific episode cost measures for mental health condition, COPD, asthma, so it's very broad kind of recommendations. But to the extent that you have experience in your own work with the organizations or other entities that you do work with, when you have ideas about

specific types of measures that would actually be a foundation for a developer organization, or some other entity to kind of pick up and have a starting place for how to fill that gap, we're certainly interested in hearing that. So, open to Lisa's additions and go on from there.

CO-CHAIR LATTS: Great. Yes, I think, you know -- and I would ask us maybe even sort of where you were going, Ashlie, to take these questions and look at these areas, but even more out-of-the-box thinking, you know. Is this the right way to go about resource-use by topic, by condition, or are there different ways that potentially could accelerate -- that would accelerate this so that we're not sitting here three years from now with maybe a one in each category, or a two, because I think that will be disheartening. And it's going to take decades for cost measures to catch up with quality measures at this rate, so let's sort of open the box and think the world, and let's see what we can come up with. So, Andrea, kick us

off.

MEMBER GELZER: So, the list of conditions, I mean, ideally, yes, it would be great to have both quality metric and a cost metric for all those conditions, something that could be used generally for that chronic condition.

That said, there are going to be things that are coming up pretty rapid-fire I think in the next couple of years from a cost perspective, and a resource-use perspective that are not sustainable for the health care system. So, just take the Sovaldi example, hepatitis C. So, I mean, if there were some kind of generic kind of a cost measure that we could then apply to hepatitis C, or if multiple sclerosis is the next one, apply to those conditions, that would be very helpful.

CO-CHAIR LATTS: Almost a plug-and-play framework --

MEMBER GELZER: Plug-and-play, exactly.

CO-CHAIR LATTS: -- where you have the processes and all you have to do is fit in the diagnoses and the conditions, and it's applicable and a very rapid turnover so that you're -- you know, because it -- you know, a hepatitis C measure would take three years to develop at this point.

MEMBER GELZER: Exactly.

CO-CHAIR LATTS: Nancy, did you want to comment on that?

MEMBER GARRETT: I just wanted to -- I don't have an answer to the general question yet, but the -- I was struck by the person and family-centered care, and I assume that's the same as patient experience. And there was a New York Times piece about a year ago that said -- that had a series of photographs, and then the question was is this a hospital or a hotel lobby? And then it would give you the pictures and you had to choose, it was a quiz. And it was hard to tell in a lot of the cases. And I tell you, if you come to

Hennepin County Medical Center, it will not feel like a hotel. We have a piano but, you know, that's about it in the lobby. So, it just strikes me that, you know, when you're looking at something like patient experience, the resources that the provider has is a really big part of that, and we don't really think about that, of the tripe aim, you know, but really make sure that that cost piece is encompassing of all the different aspects of providing care. So, it's kind of an interesting one to have on there.

CO-CHAIR LATTS: Larry?

MEMBER BECKER: So, I agree with Andrea, but I'd like to build on that; and that is that, you know, cost -- you know, the data is going to be a couple of years out of date, so we ought to keep that up to date. And we also ought to get a measure of people impacted, because not just cost but, you know, influenza or one of those things might be less expensive, but there might be lots of volume. So, I think

that's something we have to -- and we ought to keep it updated, and that might be to our earlier discussion giving direction to developers, to CMS, et cetera, where to go next. And I would also urge that we break that up by commercial plans, Medicare plans, Medicaid plans, et cetera, because I think that's also important, because if this became just a Medicare, because that's what they're interested in, then perhaps, for example, OB/GYN and pediatrics would fall to the wayside.

CO-CHAIR LATTS: Never have any measures, yes. As we saw when the initial quality development work, you know, there were no OB/GYN or peds measures for years because of that. Yes, Ashlie?

MS. WILBON: I just had a question to Andrea's point about the plug-and-play. I guess my question, my immediate response when I heard you say that was about the -- like an episode just because, you know, gathering costs around

a specific conditions varies for so many conditions on whether or not you count this cost or this cost. So, I was just trying to figure out whether or not you were suggesting that as more of like a total cost perspective, or less of the episode, or just like more of a stratification that these people have --

MEMBER GELZER: I think I'm suggesting it more as a concept --

MS. WILBON: Okay.

MEMBER GELZER: -- because some -- I mean, how to structure that? Yes, I'm not really quite sure how to do that.

MS. WILBON: Okay. That's fair enough. Thank you.

CO-CHAIR LATTS: Ariel?

MEMBER BAYEWITZ: Yes, I just -- I was wondering just in terms of differentiating this group from the episode group, so we're going after specific conditions, we're looking at resources there. I mean, you could also look at a grouper. Right? And say from an overall

grouping perspective, overall RRUs, and using the grouper to allow you to differentiate providers, or plans, or lines of business. That would be a much quicker way to do this. Right? You'd have then -- and you'd be able to roll that up at a plan level across all conditions, or you could pick X number of conditions, because these groupers map the conditions, also. Right? And then you could have then within that episodic groupings, and that would be a really easy way for a plan then to roll that up at a provider level. And you wouldn't have an issue with the Ns, you know, because you'd have a lot of conditions going across. Not that you have -- you know, could -- most of the folks here are much more knowledgeable on the scientific side than I am, and I'm -- it could be there's a really good reason why we're not doing that, but just is there a reason why we don't approach it that way?

MEMBER GELZER: And, quite honestly, as I hear him saying that, that is probably the

most expeditious way to go.

MEMBER BAYEWITZ: And it doesn't have to be -- I mean, there's a lot of episode -- you know, there's ETGs, there's MEGs, there's CAVE, there's Prometheus, I mean, there's lots of stuff out there that you could look at. Each of them have their own benefits, but all of them at the end of the day do map somehow to condition.

CO-CHAIR LATTS: And I think we had -- maybe it was even Phase One, or even in earlier groups that I've been on have talked about this, and we sort of ran into the proprietary methodology being a big part of the problem from an NCQ NQF perspective because the whole methodology has to be reproducible. I seem to recall that, some of that.

MS. WILBON: Yes.

CO-CHAIR LATTS: Are we --

MS. WILBON: We have, actually -- well, a two-fold answer, I think, to Ariel and then to Lisa. To Ariel's point,

we are -- we do actually have a separate committee that's convened around just trying to determine what it takes to actually evaluate a grouper, what that means and so forth. So, there are -- there's a CMS grouper in development, and as you mentioned there's a lot of other commercial groupers on the market. I think we're still trying to figure out exactly what that means in terms of endorsement, but I think there is definitely agreement that an episode -- for an episode-based approach to measuring cost at the condition level, that that seems to be a very amenable route to go.

We have in the past tried to evaluate kind of individual episodes that are part of a grouper. It was very difficult, which is one reason why we wanted to do this separate project around episode groupers and trying to think through what that means. Okay, if we say we're going to endorse a grouper, you know, how does that work for each individual episode, and then the measures that are used, you know. So,

it creates this cascade of kind of questions on what that means. So, it's definitely something that we're exploring as NQF. I think we're a ways away from that, probably at least a couple of years for the CMS grouper, and the value of commercial entities having an endorsed grouper I think is still under discussion, so given some of the proprietary issues and the other discussions that have been going on.

MEMBER BAYEWITZ: Yes. You know, I mean, the reality is RRUs do also take a lot of work from a plan perspective as do use measures. All these things are -- and, you know, NCQA or whoever is saying this is how we define X. There could be a way to define X around groupers. Right? Around episodic groupings, and that would be a standard way, and that way -- you know, plans may still choose to use one of these proprietary groupers for other purposes, but at least then they would have -- there would be a standard out there that they could use for measuring resource-use.

CO-CHAIR LATTS: Gene, and then Cheryl.

MEMBER NELSON: Building, perhaps, on what Ariel was saying, that there are always populations, and sub-populations, and sub-sub-populations, so it might be the State of New Hampshire and the Town of Keene, or it might be the United States, the State of New Hampshire, and the Town of Keene, or it might be everyone, and then everyone with a chronic disease, everyone with multiple chronic diseases, everyone with asthma. So, we have -- we always have this reality in front of us that there are populations and sub-populations, and we have the reality in front of us that we'd like to understand cost at all those levels. And we'd like to understand especially outcomes at all those levels. So, one point on the cost side is that, again, thinking as to what we need in the future is the direct cost of health care to the community, so getting at expenditures for health care by

the individual or by the payer behind the individual. And, also, we want indirect cost because the community is paying those indirect health-related costs, as well, for back pain or pick your condition. And, oftentimes, indirect costs are equal to or greater than the direct costs.

So, I think we'd like to have a structure on the cost side where you can go to the highest level and start to disaggregate it into important sub-populations defined by health systems, or by providers, or defined by clinically relevant populations.

The NQF Efficiency Measurement Framework, I think it came out in about 1910, or 2010 --

(Laughter)

CO-CHAIR LATTS: I was like wow, it's really ancient.

MEMBER NELSON: It seems like a long time ago, but it's pretty relevant that there's the idea of a population at risk. This

is the Efficiency Measurement Framework, and then some of those people at risk, let's say it's the State of New Hampshire, or it's all of Medicaid patients, a population, some of those people at risk have an event, and they move from at risk to an event. And then we're going to follow that event for a period of time and sometimes it turns into a chronic condition, and we're going to follow that person over time, so three big populations, at risk, an event is occurring, and then we're into follow-up.

And to be able to attribute direct and indirect health care costs for those three large populations, then to be able to further refine it by our clinically relevant sub-populations is a good idea, I think. And that framework has been developed by NQF and experts contributing to it for about four years. And I think it goes to this issue of populations, and sub-populations, and it certainly went to total direct health care costs. I've forgotten, maybe Helen remembers,

does it also include indirect health care costs, as well?

DR. BURSTIN: I don't recall, but I can pull up the picture, the chart.

MEMBER NELSON: You can't go to an employer group without being concerned with indirect costs. Right? You can't go to any civic group that's well aware and not be concerned with indirect health care costs.

CO-CHAIR LATTS: Great. And maybe we'll go to Cheryl, but I think, Gene, maybe you could comment how -- after Cheryl, how would you then in terms of specific guidance for developers and calls for measures, how would you then break that down into measures, measure requests?

MEMBER NELSON: I think there are some, probably in this page 4, the row that says "Resource Use Non-Condition Specific," it indicates that are three measures there. And one of them, I think, is a very high-level expenditures per person per year, probably not

Medicare, probably age 65, I think came from Minnesota. So, that would represent this highest level --

CO-CHAIR LATTS: Right.

MEMBER NELSON: -- health care expenditures.

CO-CHAIR LATTS: Right.

MEMBER NELSON: And then if we start to splay that out for different populations and sub-populations, we'd start to have that architecture of being able to go up to everyone and break it down in a variety of meaningful ways depending on how we'd like to define the sub-population, by health plan, by clinically relevant group, et cetera. So, starting at that level, and then carrying that down, cascading it down would be a way of dealing with it. Maybe we focus on the 20 high-impact conditions as a result of -- as a way of prioritizing clinically relevant sub-populations.

CO-CHAIR LATTS: Okay, great. Cheryl.

MEMBER DAMBERG: Yes, thank you. One of the things that I guess I've been struggling with related to the cost and resource-use measures, it seems like much of what we've been looking at has been really aggregated. And I guess I've been intrigued by some of the work that Howard Beckman has done with some health plans and medical groups, taking the information that's generated from these episode groupers and trying to deconstruct it into a set of overuse measures. To me, I think if there's some way we could signal to measure developers that we would like them to go in and try to, you know, flesh out the areas of variation and identify those areas where differential use of resources doesn't necessarily achieve greater patient benefit. I think those are going to have, you know, greater precision and greater face validity with providers.

CO-CHAIR LATTS: Great, thank you. Other comments? Nancy, did you have --

MEMBER GARRETT: So, I was just thinking about as my provider, as we're entering into more ACO arrangements, what are some of the things that we're trying to understand as new measures of cost and resource-use? And, you know, a lot of them are the classic actuarial health plan measures that health plans have been tracking for decades, but there aren't necessarily standards about how to measure them, so it's ED visits per thousand, it's in-patient visits per thousand, it's ambulatory visits per thousand. NCQA has some standards for those, I don't know that they're publicly available. I don't know that they're endorsed by NQF, I don't think they are. So, there's lots and lots of different ways of measuring those, so would that be of value to have some endorsed national measures of utilization that increasingly providers are being held accountable for and are looking at? So, that's one thought.

And then kind of related to that,

some of the other -- I mean, it might -- a place to go might be to just at some ACOs and what are they tracking? So, another measure that we're really tracking is the percentage of ambulatory care that's in specialty versus primary care. And that's something that varies quite a bit across providers, and really is a huge cost driver. If you move more of that care back into primary care for a whole population, you can really save a lot of money and do things more efficiency, but it really has a lot to do with practice patterns, but it's very amenable to intervention. So, that's kind of some more -- I think we might find some more creative ideas for cost and resource-use if we look at what the ACOs are tracking and what they're starting to look at.

CO-CHAIR LATTS: So, I virtually put my own placard up. I think what is really occurring to me throughout this conversation is that we've sort of done the shotgun approach up until now, you know. We need all these

measures so we're going to have these various phases and bring stuff into us, and we'll review it. And it almost makes me wonder whether or not NQF should seek funding because, of course, it's all funding-generated for a exploratory group, committee, you know, smaller than this, you know, eight to ten people or something like that, that would then go out, gather those measures and prioritize a list, and then bring the prioritized list to a Standing Committee to endorse, or suggest, or give feedback on, and then go out and seek -- you know, request measures based on that prioritized list rather than, you know, sort of this shotgun approach that we've been doing up until now. Helen, I'll let you --

DR. BURSTIN: Well, that was timely. I would have responded to your virtual card being up, anyway. But, quickly, I just wonder -- just to Nancy's point, which I think is a really interesting one about utilization measures, it's come up a lot over the years. And

one of the concerns when they've been raised is should we bring in utilization measures, because by themselves they have no quality signal? Well, maybe we cross that bridge by bringing in cost and resource-use measures which we know also in and of themselves don't have a quality signal. And maybe it is time.

I mean, one of the issues that comes up a lot as we try to align measures, for example, with the private health plans is they have lots of utilization measures. I assume purchasers like you, Carolyn, use utilization measures all the time. So, that -- but, actually, that's an interesting approach, and there's lots of those out there, and it might be useful to get a prioritized list of which ones. ED visits, for example, comes up constantly, and every time we've tried to review it, this whole issue of avoidable and preventable, and all the stuff attached to that gets really complex, as Brent knows all too well. But maybe as building blocks, just

starting to think that through, it might be interesting.

But to Lisa's point, which is a great one, that's actually been a lot of our thinking about this idea of trying to move to a measure incubator, trying to bring all the right players together saying here's the big ideas, this is how they're prioritized, here are the -- here's potential funding for it, here are the experts, here are the developers. Can you take this measure used in Minnesota and help create it into something that could be a national standard? So, we've had lots of discussions like that, and one of the key issues we have to figure out, as well, is how do we find out what's being used on the ground? I think that's -- we have a good sense of it through federal programs, you have a reasonably good sense now with some of our work with Michael Bayliss' group in terms of at least 25 states' measures, but there's so much more, and it just -- again, we don't have -- the IOM talked

about having this continuous-learning health care system. And without those kind of feedback loops or information from the ground, we don't have it, so your thoughts on how to get those kind of data would be really useful.

CO-CHAIR LATTS: Brent.

CO-CHAIR ASPIN: To Nancy's point on this thread, that's what ACOs are looking at right now is utilization, because we don't -- I mean, we don't have timely enough data. And even that's challenging because you get out of your clinical EMR and you have to use claims in order to have reliable information about it. Right? But, yes, I think that would be a fruitful area for looking at resource-use, and there would be a number of different areas where you could tie it to the quality framework that's in place.

MEMBER GARRETT: There aren't standards, so as opposed to a harmonization, it would be really useful.

MEMBER BECKER: And maybe this is sort of opening up another box, but I started

to read this and I said so, how do we begin to engage or put in all of this what the provider and the patient we've talked about and come up with a preference around some option? Because if I read this right, it says on page 7 from the 9th line it says, "Requirement that patient receive all recommended care for the composite to be met."

Well, some patients might not prefer all of that care. A cancer patient might decide to go all in, or not. A hip replacement on a skier might go one way, and a person in their 70s might go another way, so how do we begin to sort of bake in patient preferences into all this measurement?

CO-CHAIR LATTS: So, actually, Larry, that gets into one of my favorite topics which I've been talking about for several years now, which is that as we start to get sophisticated in our quality measurement and now increasingly in our cost measurement, where does that come up against the whole idea of

patient empowerment? Because, you know, in patient empowerment the idea is we're going to let patients make informed decisions based on good information, and based on what's relevant to them. And if that means that I'm not going to get this or that recommended test, and maybe it's going to cost less, maybe it's going to cost more, that that's my decision as a patient, I think we're in for some interesting times.

MEMBER BECKER: So, a comment on this as a patient and ask people who are providers to comment, as well. So, as we begin to put payment at risk here, there's this conversation between the patient and the provider, and really trying to come to some agreement as to where we're going to go. So, when you have that dyad and you have these measures and those pressures, how do we manage through that?

MEMBER WALKER: From a consumer perspective, a lot of the cost and resource-use measures we've looked at recently is just -- I'm sorry. A lot of the cost and resource-use

measures we've looked at recently are not very helpful to consumers. You know, most of them are more interested in what their out-of-pocket hit is going to be, and we don't -- I mean, we've not considered that at all.

That said, that doesn't suggest that these measures are unimportant. I think that they're really important. You know, I'm -- Jennifer isn't here, but I'm channeling Jennifer, and one of the comments she made at our very first meeting was that maybe a developer, or maybe CMS isn't necessarily responsible for interpreting and presenting the information in a way that's used for consumers, maybe consumer groups can do that, too. But I just want to put it out there that, I mean, I think -- developing these measures I think is one step towards getting into information -- creating information that we use for consumers. In and of itself right now we're not there yet. I think that we continue along this road, we'll get to a place where consumers

will be able to have the tools they need to make more informed decisions.

I mean, these days a lot of employers are asking consumers to take on more accountability for their choices, and lots of places are moving towards consumer-driven health plans, or high-deductible health plans where they have to assume those responsibilities and accountability. So, having good quality and cost measures will be very critical, not just for providers and for plans and hospitals, but also for consumers. So, you know, I don't know -- right now I can't offer what exactly those measures are, but I think with more thinking and conversation we'll get there.

CO-CHAIR LATTS: Good. So, where does this leave us? You know, let's answer that question. Yes, I mean, in terms of the questions that NQF asked us to review, you know, where are we? Brent.

CO-CHAIR ASPLIN: I don't know if I

can answer where we are. I mean, there's a lot of good news/bad news in a lot of these conversations. At a macro level, I think we are trying to skate towards where we want the -- we think the puck is going to be. Right? Think about the measures we're reviewing in this context. We have a combination of longitudinal total cost of care measures over a year, and episode-based measures that are designed to be tied to quality measures. And I agree with your point that maybe it's not as consumer-relevant now, but I think those are the two right buckets to be in, you know, bundles and global payment and quality, and total cost of care measures.

You know, how we translate that more to relevance to consumers, I think the comment was made earlier today that we maybe have to tackle the standardized pricing piece at some level if it's really going to be compelling. Now, that creates potentially enormous complexity but, you know, if you think about all these measures outside of Medicare, that is

probably one of the most important missing links both in terms of out-of-pocket and total cost of care differences where we would take out standardized pricing and bring in actual prices.

I don't know if that is somewhere we want to go, and what kind of methodological and market-based, and other issues that creates. It feels like a lot of the secrecy is going to get blown up anyway, so maybe it won't be that threatening of a ground to be in.

CO-CHAIR LATTS: Yes.

MEMBER GARRETT: But I thought the Health Partners total cost of care measure has one version with standardized pricing and one version with actual prices. Isn't that right?

MS. WILBON: That's correct.

MEMBER GARRETT: So, that measure portfolio actually has real prices in it. I think it's the only one. Right? So, there's a precedent that ---

CO-CHAIR ASPLIN: Well, I don't know

how Minnesota Community Measurement is --- so, they were planning to take this to the next level with transparency on total cost of care, and I just don't know which version they were using. I know Health Partners shares your --- both our resource-use index and a price index for each condition suite when they sit down and talk with you, your delivery system in their market, because I've had those conversations.

Now, I don't know which versions, and Jim or Nancy, maybe the two of you know which version Minnesota Community Measurement --- which track they're taking, because all I know is that what's in process, but I haven't gotten an update since I left.

MEMBER GARRETT: I was just saying that the endorsed NQF measure does include real prices, so if anyone is using it as it was endorsed they would be showing real prices. And the Community Measurement more work is in process, I do believe they're going to be

showing both versions. I believe the plan in Minnesota is to have both versions.

CO-CHAIR LATTS: Real and standardized.

MEMBER GARRETT: Yes.

CO-CHAIR LATTS: Which, you know, is -- I think gives you the best of both worlds because you can look at actual if that matters, but you can also compare apples to apples for a standardized methodology. I think Gene was next.

MEMBER NELSON: Maybe building on Brent's comment of total per capita costs and bundled costs for certain kinds of people, certain kinds of situations, and where we need to be going where the puck is. So, this image may not be right but I like it a lot. If I'm a patient now and if I go to the Spine Center at Dartmouth, I can see --- and let's say I have a herniated disk or a degenerative spine. Based on data collected at Dartmouth and 13 other

centers, I can look at my calculator and I can get a sense of my disease burden at one, two, four, and six years for my back problem. I can get a sense of what my physical functional health will be using, for example, an SF-36 physical. I can get a sense of for the average person like me, if I got surgery or if I was treated non-surgically my satisfaction with the treatment benefit that I got, and I can look at my direct and indirect costs of care in a rough way. And then I could go to our payment office, basically, and they would say well, Gene, since you are on a Medicare or since you're on Harvard Pilgrim, this is what your out-of-pocket costs are going to be related to the spine treatment approach you pick for the next 12 months. That's pretty powerful information. That's what I would like to have.

CO-CHAIR LATTS: Yes. Ariel.

MEMBER BAYEWITZ: I was just going to say from --- I mean, the comment was made around consumers. I think in thinking about consumer

perspective, RRUs are probably confusing, and it doesn't really get to where they want to be, which is what is it going to cost me. And I think you need costs.

Now, we know what the problems are there, right, specifically around provider contracts. And I agree, at some point, probably in the near future that we're probably going to get through that just because of exchange membership, and all of these new consumer-directed benefit programs. They're just growing rapidly.

But I do think if we're trying to think about the consumer, you need to know costs. And even in the example Gene mentioned, I mean, the person is going to want to know what is it going to cost me, not what would it cost me if there was some normalized fee schedule.

I think where it is helpful, and this is not always the same as a bundle, but sometimes knowing cost for service can be confusing, also, for a consumer because, you

know, you get a procedure done, and there's the hospital cost, and the pre-op cost, and the physician versus the hospital, and there's the lab, and the radiology, and then the --- even within the professional provider the anesthesiologist, and there's a pricing there. It's per 15-minute increments, and other persons per visit. I mean, it's very confusing, so to just show cost per service I don't think is helpful. There is a value of having one way to sort of summarize --- synthesize that for a consumer to say this is the standard way that we look at endoscopies, for example. And this is what costs here versus here.

Again, we have this problem with people sharing that information, but I think if we're trying to think about a consumer, that would be helpful.

CO-CHAIR LATTS: So, I actually raised my placard that time. As this is proceeding it seems to me we're trying to boil the ocean. I don't know that we can be all things

to all people, and I don't know that we can develop measures for consumers, and measures for health plans, and measures for providers, and the public, and the private.

I'm just not, you know --- I just don't know how I see that happening long term. But also reflecting back onto Andrea's earlier comment about the --- you know, we need a generic template that we can just plug-and-play our top 10 conditions, or be flexible. I mean, it seems like we --- you know, it's not perfect, and we've discussed it, but the measures that we've reviewed at these last two meetings are standardized measures. I mean, you know, the RRU measure and the 30-day cost measure, you know, have essentially been plug-and-play for three different conditions now. And, I mean, I think that's at least a guide for something potentially going forward that, you know, I'm sure it's not cheap or easy but could be at least somewhat rapidly adopted rather than starting from scratch.

CO-CHAIR ASPLIN: Backing up for a second coming at this focusing on the per capita longitudinal cost side of the equation. You know, just looking at the ACO 33, I mean, it's in four domains that make a reasonably compelling case when coupled with the cost performance against a target from a purchaser's perspective. You know, you have patient experience, you have at-risk populations, you have care coordination and safety, you have prevention, and that suite is reasonably compelling when coupled with your cost performance to achieve those outcomes.

Now, there's lots of details, and we could spend a long time talking about the challenges with the specific measures, and attribution, and so forth. And I still don't think it gets to your comment, Lina, around consumer engagement, although, in the context of the exchange that might begin to evolve in that direction. I think it's maybe some measure of engagement that needs to get brought into the

domain on CGCAHPS and other experience measures. That might be one of the missing links there, although that in itself is still probably going to be compelling to purchasers and employers than it is to individual consumers.

The point I'm trying to get at here in maybe too long-winded of a way is that we had a bite at that apple. Right? And we passed on it for lots of good reasons; don't take that as an editorial. I'm just saying we had the total per capita cost into the suite and it feels like at least in the Medicare space, you know, we're going to have to kind of --- part of the conversation should be how do we come back at that? Most of our concerns are around the attribution model. There were a list of others, risk-adjustment and attribution, common themes. Right? But a balanced portfolio in the NQF space of some subset or, either validation or addition, or comment on the categories of quality experience and so forth that we have in

the ACO 33 coupled with something that we could get through this process that we believe in. It's got to meet the criteria on total per capita cost I think would be really helpful for those that are in the longitudinal space. We kind of have that in commercial but we're missing it on the Medicare side.

CO-CHAIR LATTS: Carolyn.

MEMBER PARE: I was going to make comments very similar to yours, Lisa, in that I don't feel a lot of the work that we're doing right now here at the Quality Forum resonates with consumers, and I think we're a long way away from it. And I sometimes am concerned even with some of the conversations we have right now about total cost of care and SES that we're going to make it even more opaque for consumers because we need to satisfy providers that things are fair. And I don't know that people --- so, I don't know if I've shared my story in this group before, so throw something at me if I have and you're bored with it.

But when we rolled out exchanges in our state, there were some markets, and this happened to me because I walked into a pharmacy and I was getting a prescription filled, and that's one story just in the prescription-filling piece, but in the other piece the way it summed up was the pharmacist tech said to me, you think that's bad. My parents are down in New Ulm and they pay here in the Twin Cities, they could get their insurance for $200 a month, but in New Ulm it's $500 a month, and that's because of that damned Obamacare. So, I had to say, well, no, it's because there's a large health care system in that part of the state that really has a monopoly and can drive the prices up. And that's something that the basic consumer doesn't understand, so basically when costs are high they'll blame the health plan. They'll talk about all these things that they've heard or seen in the media about those people being on the take. They really don't understand all the

different cost drivers, and so much of the work we do here is about the cost drivers, the quality that drives poor cost results.

CO-CHAIR LATTS: Matt.

MEMBER McHUGH: It just reminded me of a conversation that we had, and a reminder, and I'll channel my inner Jack Needleman that this is where we're focusing on billed services. And there's a whole range of providers who are invisible to that even though their variation in provision of care certainly will affect quality, but also gets absorbed in these measures around cost.

CO-CHAIR LATTS: Brent.

CO-CHAIR ASPLIN: I'll just make this quick. One other thing just to throw in the hopper that we capture it, is trend. Okay? So, we've talked a fair amount about acuity-adjusted costs in a given performance year, and we have not yet really brought in trend into the conversation.

Again, I don't think this will be

compelling to consumers right out of the gate. I really do think over time it will become more and more compelling to consumers and will help address some of the concerns that organizations that are taking on risk are worried about in hopping from year, to year, to year just over minuscule changes in price on a premium, that if measures that have made it through this process could become reliable rival measures of how well combinations of delivery systems and financing organizations can hold down costs over time, I think that would be compelling.

CO-CHAIR LATTS: Great. Sorry, I'm not sure who was first.

MEMBER BECKER: I was going to say, so I agree that consumers and most people don't really understand the hydraulics of health care. I think what we have to do is challenge ourselves to think about these measures, who's the customer for the measure, and who's the end-user of that measure? And how do we construct it so that those two parties can

actually use it? I think we've got to challenge ourselves.

CO-CHAIR LATTS: But to that point, and to many --- I mean, do consumers really even care about the cost of health care? They care about their cost of health care.

MEMBER BECKER: That's right, they do.

CO-CHAIR LATTS: Yes.

MEMBER BECKER: They want to know their price.

CO-CHAIR LATTS: Right.

MEMBER BECKER: You know, what it costs them. And so, we be careful when we say the price of health care.

CO-CHAIR LATTS: Right.

MEMBER BECKER: What we really mean, will it ---

CO-CHAIR LATTS: Is what does it cost me? Yes.

MEMBER BECKER: That's right. What does it cost me? And oh, by the way, the

providers are caring more and more about providing the patient with the cost that it's going to --- they need to extract from them ---

CO-CHAIR LATTS: Because they ---

MEMBER BECKER: --- at the visit because ---

CO-CHAIR LATTS: Bad debt.

MEMBER BECKER: You know, balanced billing, and bad debt, and all of those costs are rising dramatically.

CO-CHAIR LATTS: All right. John, then --- oh ---

MEMBER DAMBERG: This is Cheryl. I want to concur, I thought it was Larry speaking.

CO-CHAIR LATTS: It was.

MEMBER DAMBERG: Because I think with very specific end-users just for the use measures, and that purchasers are acting on behalf of patients in trying to reduce health care spending so that it is more affordable. So, I don't have any expectation that the measures we adopt through this process will be

ultimately useful by a consumer, but I think we have agents who act on their behalf.

MEMBER WALKER: And that was exactly my point. I think a lot of the work that we're doing is to make pricing --- well, maybe moving towards making pricing, but definitely cost and quality more transparent to consumers. And if purchasers are able to make better decisions, ultimately, it benefits consumers, so definitely agree with what Larry and Cheryl just said.

CO-CHAIR LATTS: Right. John.

MEMBER RATLIFF: Although the end-users of our measures are not consumers, they're going to be the plans that are utilizing the quality metrics, or the resource-use measures, other things that are being vetted by NQF, so the NQF work product is really not something relevant to consumers.

I also offer that consumers of health care probably also have some idea of value, too, and if you can provide them data as

to what the return on their investment is going to be, then they may be not so concerned about the out-of-pocket expense but what they're going to get for it, such as what Dartmouth is able to do with the Sport Study, with the data they've wisely provided.

And, thirdly, this really reminds me of the conversations we had on the Overuse Committee like seven or eight years ago when we were talking about patient autonomy and choosing to have a stress test, or choosing to have a cardiac intervention.

I don't know that NQF can answer those concerns. I remember being on that panel probably seven or eight years ago and I do recall at that point a patient would come into an ER with a headache and want a CT scan. And if he didn't get a CT scan, he'd be unhappy, but thanks to Choosing Wisely, and thanks to a very concerted effort to educate patients as to the radiation exposure entailed by CT scans, now when I as a spine surgeon order a CT scan, I have

to talk the patient into getting it done because they're terrified of the radiation exposure. So, there is a way to get around this with regards to educating consumers, patients. But, again, I don't know if that's within the scope of NQF. I don't know if we can do that.

CO-CHAIR LATTS: Yes, good point. Nancy.

MEMBER GARRETT: Just following up on John and Larry's point about what consumers care about. One thing to keep in mind is that when you're talking about costs and health care, you know, you're trying to make a complex consumer decision, and consumers don't necessarily want the lowest price option because there's conflation of price and value. Just like with a lot of things we buy, do we want the Cadillac surgery, or do we want the really cheap car that's going to break in a month? So, that --- you know, until we have really, really strong measures of quality that completely tell the story, that's going to be conflated, so I

just want to throw that out.

But then, Brent, I had a question for you about the trends. So, I'm not quite sure I understand what you mean when you say we need some endorsed measures, ways of measuring trend. Is it the --- like if you're trying to compare an ACO's performance from year one to year two, is it what you're building in for what would have happened naturally in environments so that you could understand the impact that you've had, or is it --- what exactly ---

CO-CHAIR ASPLIN: No, I think it's just a matter --- I don't think it is necessarily new measures, but just starting to expand the viewpoint over a period of time. You know, like for example, the NCQA measures we had today were not constructed in a way, we had that conversation, that would enable that, the per capita total cost of care. Yes, they're shifting populations to some degree, and we have to sort through those details to understand that you're not just changing your

mix in order to make it look good. But, you know, the question is why should I make a commitment to you over the long run? That's the question I would propose that would be an understanding space to try to answer over time.

CO-CHAIR LATTS: Carolyn.

MEMBER PARE: My earlier comments were never to imply that the work we're doing here has no value. I just ---

(Simultaneous speaking.)

MEMBER PARE: --- which I suddenly felt. The reality is yes, we do want to engage consumers, but I don't know that that's going to happen in the short term because of just the way the system is constructed, and the work that we need to do in order to get there.

I do agree with the other purchasers and consumers in this room that one of the reasons NCQA metrics are so important, even though they just measure at the health plan level, is that's typically the organization that we look to in order to be our stewards of

our cost. So, a lot of employers work through their health plans in order to manage the providers underneath that within the system so it's important work.

I just need to keep on coming back to it, because sometimes we get very circular in our discussion, and we do get to do exactly what you said, Lisa, and that is we boil the ocean. We're trying to think about how these measures will work for everyone, and the answer is they won't work for everyone. There are different measures for different stakeholders, and we have to hope that they resonate with those stakeholders that can move the market the best.

CO-CHAIR LATTS: Ariel.

MEMBER BAYEWITZ: Yes, just on the --- two points. One, on the consumer price piece, you know, I would just echo what a lot of folks have said. I just add that benefit differences for the member, those benefits vary widely by plan, within plan by product, within

product, there's employer groups that have different benefits than others, so it gets very tricky if you're trying to have sort of one approach to how do we show cost to a consumer, because cost --- it's not even --- even if there was one rate to the consumer, that would manifest itself differently.

Just to Larry's point, you know, I think this is what I was getting to earlier. If we had to prioritize measures, I think it is helpful to think about who is the end-user, who is going to take action on this? And it could that be that we could say, you know, for resource-use, for example, I think providers own those --- providers determine which resources are used, you know. It's not to say that plans aren't involved. Plans could incentivize providers to make certain decisions. And that would be a good example where you could actually capture something at a plan level, but you need to make sure that whatever you're measuring is also meaningful at

a provider level if you expect that providers are the ones that are responsible. You know, give you an example, you know, as part of one of our big programs we're looking at well, which utilization measures should we use to evaluate, you know, some outcomes? And, you know, a lot of people mentioned well, why don't you look at these certain readmission measures? They're endorsed by, you know, lots of folks, and we said well, we're not going to use that because at the end of the day it's not statistically meaningful at a group level. You can have a 20-30,000 member group level, some of these readmission measures just aren't meaningful, so we're not going to build a program around that.

You know, you could look at potentially preventable ER visits, ambulatory-sensitive conditions, there are measures that at a group level even not a huge ACO, but even a large provider organization that is meaningful. And if you would evaluate

plans measures like that so then they would understand at a plan level how are they performing against other plans, but then they could actually take those same measures and incorporate it into actionable programs and incentivize providers. So, if --- and I don't know what that ocean is of measures out there. If we could think about, to the point, who is the end-user, who's going to take action, use that to then drive prioritization, obviously, and keeping in mind cost, you know, overall spend out there. You're not going to go after a measure that doesn't have an impact, but I think that would be helpful.

CO-CHAIR LATTS: So, calling on myself next. I sort of wanted to emphasize something I've heard here that I think is really interesting, which is almost a return to simplicity: the idea that every health plan and probably most of the larger provider groups out there are using basic level --- basic assessments of utilization, and cost, and

resource-use that are used every day, every quarter to assess every single plan=s success, that now with the rise of value-based purchasing and ACOs and provider-based measures to have those standardized and endorsed might provide some value. So, almost saying okay, let's step back from the complexity of what's on whatever table this is, and let's do basic utilization measures, and standardize those. And I don't even know who would bring them forward, or how that would work, but that that might, indeed, have some value in today's environment that wouldn't necessarily be valuable five years ago.

MEMBER GELZER: Lisa, would you clarify those measures? You're talking about days per thousand.

CO-CHAIR LATTS: Yes, the basic --- you know, days per thousand trend numbers, PMPM costs for inpatient/outpatient pharmacy professionals, you know, some of the basic things that health plans do as bread and butter

that are now --- you know, provider groups and others will be held accountable through value-based purchasing.

MEMBER BECKER: So, this is a really good conversation. And going through my mind is okay, so what do we do with it? And would it be within our purview to set up with a multi-stakeholder group a series of I'll call them principles about where we want to go, get buy-in from all the stakeholders, and put that out there as sort of this is sort of directionally how all the stakeholders think we ought to move going forward. Put it out for public comment, get that kind of input so that we can set some kind of agenda, doesn't have any, you know, teeth in it, right? We can't enforce it, but at least it's a considered approach by all the stakeholders.

CO-CHAIR LATTS: Does NQF want to comment on that?

DR. BURSTIN: That's part of what you're doing right here. You have the

stakeholders at the table so I think some of this is --- put out what you're going to come out of this, and maybe we can invite further dialogue, Larry, I think about the path forward and get some input on it and the path forward of how you actually make it happen.

MS. WILBON: I'll just add, I think there are some things that --- at least I've been taking notes here, and there are some things that I think we could kind of turn into principles if that's the direction that the Committee would like to kind of frame it in. I think that's definitely doable in terms of us capturing the discussion and putting it out for comment.

CO-CHAIR LATTS: What I really like about that idea, Larry, is I think it would help us --- it would keep us from boiling the ocean. If we stick to our principles and prioritize based on the principles it might give us some direction that otherwise we're flailing about. So, I didn't see the exact order, so I think Jim,

then Brent, then Ariel. Okay.

MEMBER NAESSENS: So, kind of reflecting back on one of the things Brent said is that we really have kind of two approaches in terms of our cost and utilization measures. And as a provider organization we really have two different types of patients, we have, in essence, the capitated approach where we're really providing primary care and community care, capturing all of those capitation bases. Then we have kind of, in our case kind of a destination patient or a referral patient who's coming to us for a short term, short period of time. And, unfortunately, a lot of the population measures don't apply very well to that group of patients, so we need to keep in mind that to be in an effective --- to come up with effective measures for our sort of practice we need to have almost a two-tiered approach.

The episodes seem to do a reasonable job with the destination patients for those

kind of patients coming to us for elective surgeries, things like this. And the capitated population model works very well for the primary care model, but we can't always C-- and measures currently don't always separate those two groups very well.

CO-CHAIR LATTS: Ariel, Brent is going to switch, so go ahead.

MEMBER BAYEWITZ: Yes, just on the comment of looking at days per thousand, claims per thousand, visits per thousand, those kind of things. So, I guess I would wonder what --- if we would look at a measure like that, who in that case is the audience because the --- each payer probably has --- I mean, I know they have their own method of calculating these things. And I don't know how meaningful it would be to compare payers on measures like that. You know, I don't even know, you know, like a days per thousand, honestly, a lot of that stuff is driven by contracting and how the rates are structured. Are you paying per day, are you paying per base,

and that has a huge impact. And it's not always necessarily true that it's bad to have a few more days in some cases, so I would personally be interested in looking at not just overall days per thousand, or overall admissions per thousand, or overall ER, but more targeted like potentially avoidable ER, like ambulatory-sensitive conditions per thousand which are more specific and people would argue much more actionable than just a more general, so I don't know if there's interest in that.

But the other piece on days per thousand just blanket ER visits, you know, I mean, I know with our states there's different ways that people code in different markets. And how you, you know, for ER, do you include a 456 revenue code as ER, or urgent care? I mean, there's little things like that that actually are market-specific so, you know, something to think about.

CO-CHAIR LATTS: It occurs to me, though, that the place for the standard or the

approved measure of those things would be not within an organization, because you've got your thing, you do your thing, but between organizations. So, where there's a value-based purchasing agreement so that a provider group doesn't have to deal with the five different methodologies of each health plan they're contracting with, or in cases of transparency where a division of insurance, or a state government is requiring a health plan or provider group to make certain information public, to give them a standardized way to require that. That's what I was sort of thinking.

MEMBER BAYEWITZ: Yes. No, I agree. I mean, we --- and that's how we use HEDIS right now for quality. We say, you know, we know it's not perfect but it's been sort of approved to some extent by everyone. It's not perfect but it's something that --- it's a common measure that everyone can use across payers and across providers.

CO-CHAIR LATTS: Apples to apples.

MEMBER BAYEWITZ: So, how do we get a similar kind of thing for targeted utilization measures?

CO-CHAIR LATTS: Right.

CO-CHAIR ASPLIN: Yes, I think --- just to pull the thread on this, and then I want to open a different topic, maybe. I think the audience would be --- would follow an attribution model. I think it's really an accountability for primary care teams which is where we would want most of the utilization for a population, not an individual patient, but a population, most of the accountability to reside there. I mean, that's the construct in ACOs, and I think that would be helpful.

What would really be helpful in addition to the fully adjudicated final answer would be if there are ways to get early warning signs of utilization patterns in real time, because one of the challenges is we get this

really clear detailed depiction of the accident scene, you know, a year later, and it doesn't help at all at that point because other than tell the story, and then you try to change some things you find out a year later. Now, what those of us in most organizations are starting with their own associates, their own employees, they're trying to do it, and you have your own TPA function, is try to get early signals well before the fully adjudicated claims run-out period so that I can understand, you know, when so and so is in the emergency department more in real time that I'd have the opportunity to intervene and prevent the bounce-back ED visit, and then the third ED visit that results in the hospitalization. That would be a full other area of complexity, but it would be interesting.

Reflecting on this list, if I just shift for one second, I would be interested, not necessarily right now, but I would be interested in learning more about the 63

quality measures in health and well being, because if you think where we want to be 10, 15, 20 years down the road, I think we've made a lot of progress starting at the intensive utilization end of the spectrum, and for good reason. Right? That's where the money is. It's sort of the Willie Sutton approach to how we understand cost and resource-use. And right now those are the people that are at greater risk for harm due to their care.

Over time, though, we obviously want to keep the --- you know, narrow the pipeline of people getting into the categories with multiple chronic conditions, and it would be very interesting to understand, particularly since health care only accounts for about 15 percent of health status, anyway. Right? Forty or 50 of it is health behaviors, 20 genetics, 20 environment. To the extent that some of these 63 measures are around health behaviors, it would be interesting to segment the populations by band, and then across into

different buckets starting with the whole category of people who have no chronic conditions. And what's the cost to achieve high performance in certain measures of health behaviors and wellness, because it's really around engagement, engaging people in maintaining their health, and maintaining healthy behaviors. And I don't know what the right answer is. Should the cost to do that be low or high? Maybe we should be investing more, so we've had similar questions in that regard. But we really need to start thinking from the other end of the spectrum in addition to resolving some of the complexities in the current space we're in.

CO-CHAIR LATTS: I was thinking about that. Gene.

MEMBER NELSON: So, once again building on what Brent just said and thinking about the future, something --- measures that would apply to larger populations that can then be cascaded down. I think there's three major

buckets of measures of health outcomes, and this to some degree relates back to that earlier comment on the NQF Efficiency Measurement Framework where we have three populations. So, the big three measures might be risk status, avoidable risk of death, that can apply to a population or to individuals, so think of Framingham-plus, avoidable risk of death. That's a population one.

The second big category of measures is disease burden. And now that I'm out of the risk group and I have one or more conditions, what's my burden of disease? And this especially goes to the multi-morbidity person, but if I just have bad asthma, I've got a disease burden. How does that get measured and improved?

And then where people live is the third. It's their everyday functioning, physical health, mental health, cognitive function, you know, ability to do what I need to do. And that, especially, goes to that third

group where you're living with different problems. And there's the doctor's problems, and that's the high blood pressure, and the bad blood sugar, and all this other stuff, but there's my problem, and it's getting around and feeling mentally acute, and not being depressed, et cetera, physical health, mental health, all gotten very well with modern measures like PROMIS.

So, there are three big buckets that apply to people, risk status, and that generally gives way over time to disease status and states, and that gives way to limitations in physical functioning, mental functioning, et cetera. And just as we would like to have global costs that can be cascaded down, we'd like to, I think, have risk, disease and function that can be cascaded down.

CO-CHAIR LATTS: It's overwhelming to me just thinking about it. It seems hard.

MEMBER NELSON: It seems hard, but there's a lot of good work that's been done, and

it's increasingly possible.

CO-CHAIR ASPLIN: If we want to buy health rather than care.

MEMBER NELSON: Right.

CO-CHAIR ASPLIN: That's kind of where we need it.

MEMBER NELSON: Right. Yes, we're -- some places are in a pretty intense dialogue around health and health care in a region, so now how do we get best health and lowest cost health care in the region? And you start to have to go into the health care and the total cost and the determinants of health. And that, you know, is part of our future. It's not very pressing today in most places, but the Aligning Forces work at Robert Wood Johnson Foundation, the Rethink Health work, many of these are starting to bridge those areas.

CO-CHAIR LATTS: It's important. It feels --- this is one of those big, hairy, audacious goals. All right. Anybody else want to have the last word? Otherwise, we'll give it

to Gene and move on to the measure. Oh, we're actually scheduled for a break right now, so we'll have a break and then move on to the measure. Sorry, guys. Anybody on the phone want to make any comments?

(No response.)

CO-CHAIR LATTS: All right. Then why don't we take --- oh.

MEMBER BECKER: Do we have next steps now? I mean, we've had this conversation before ---

CO-CHAIR LATTS: We have it tomorrow.

MEMBER BECKER: Okay.

CO-CHAIR LATTS: Next step is to talk more tomorrow. So, the good news is we have cut some time off our day tomorrow because of having the conversation today.

Okay, great. So, let's take a 10-minute break. Yes, 2:45 back here, and then we'll do the pneumonia measure.

(Whereupon, the above-entitled

matter went off the record at 2:31 p.m. and resumed at 2:45 p.m.)

CO-CHAIR LATTS: Welcome back and welcome to our measure developers. So, we are about to embark on our third measure of the day so we're really cranking today. And this is the episode of care for pneumonia. And as you all recall, we had our discussion earlier today going over the Phase 2 measures, and our concerns with that. Do these guys know about the revotes and all that?

MS. WILBON: I believe, yes, you guys got the voting results for Phase 2. Correct? Yes, okay.

CO-CHAIR LATTS: So, that there was some changes in certain members' votes. And the --- it was related to changes in the understanding of how the measure was developed and the history behind it, and how applicability, so that was some of the reasons for why the votes were changed, and a better understanding of measure use.

So, I think that there was --- yes, we're talking about Phase 2, yes, the AMI measure. Yes, Phase 2. And I think one of the things we also noticed was that you addressed some of our concerns in the AMI measure in the packets for the pneumonia measure, so we do appreciate that.

So, with that said, I'll hand it over to Brent, and we'll start talking about pneumonia.

CO-CHAIR ASPLIN: Thank you, Lisa. So, we're on Measure 2579, episode of care for pneumonia. And we will turn it over to our developers. You want to introduce yourselves, again, and welcome back. And we'll just hear kind of a high-level overview briefly of the measure.

MS. KIM: Hi, my name is Nancy Kim. Thank you for having me today. Yale really appreciates the opportunity to be here. I'm a general internist trained in health services research, also a hospitalist actively seeing

patients and I led the development of all of three measures, including this one that I'll present today. But I wanted to take a moment to --- for Lein to introduce herself.

MS. HAN: Hi, I'm Lein Han from CMS. So, I would like to say a few words from the CMS perspective on payment measure, in general, and I will ask Nancy to present a summarized payment measure, a pneumonia payment measure. Okay. Thank you for the opportunity for us to present the cost measures.

CMS has developed and implemented quite a few outcome measures, and for our Quality and Payment Program, and this is first CMS attempt to examine value of care by combining quality measure with the cost measures.

These measures, when I say these it's because I'm taking into account AMI, heart failure, and pneumonia. These payment measures are not perfect measures, but they are good measures as recommended for endorsement by the

Committee. For us, these measures are the first step for moving forward. And providers and consumers, so far they have no information on cost data across care settings.

We heard in the past that people were saying that well, CMS is data-rich but information-poor. And we have a lot of data collect --- a lot of data submitted by providers. And these data are collected using taxpayers= money, so we thought we should use these data to do something good to benefit of the patients.

I want to point out that only CMS can provide this type of bigger picture and detailed cost data across settings nationwide. This measure can use --- could use the data CMS has, and the data, when I say data not just the measures, because when we implement measures we do provide a lot of data for the hospitals. So, I think this measure will good use the CMS data and the data will be useful for the hospitals.

This is a new type of measure. We

heard the debates in the panel discussion. As you are quite aware that we run into the same difficulties and skepticisms when we first started with reporting the mortality and readmission measures, and we learned that if we bring these data/information to light, and in the long run would benefit the --- would help improve the quality of care. And this is the lesson we learned.

And, also, we believe that this is an opportunity for CMS, of course, with your help to lead and begin a conversation about this type of measures. It's a measure -- it's an area unmeasured, but critical --- it's a critical dimension to health care. Thank you.

MS. KIM: Sorry, it's Nancy Kim again. Just a very brief overview of the pneumonia episode of care payment measure. I know that this methodology is familiar to the Steering Committee, so please stop me in the middle if you have questions, but the methodology is very, very similar to our AMI and

heart failure episode of care payment methodology.

So, the goal is to measure a hospital level payment through an episode of care that begins with the pneumonia hospitalization and ends 30 days post-admission. To create a relative hospital measure that reflects both differences in inpatient and post-discharge care we remove payment adjustments that are unrelated to clinical care, such as geographic considerations like cost of living and wage index, as well as policy adjustments like indirect medical education and DSH, Disproportionate Share payments that CMS makes for other reasons. We also risk-adjust for the patient case mix, and we are aligned with our publicly reported NQF-endorsed pneumonia mortality measure.

We create these measures with the chronic condition data warehouse or CCW data. These are Medicare fee-for-service

administrative claims data that contain 100 percent of patients with a primary discharge diagnosis of pneumonia. And these data include payments for the index admission up to seven other post-discharge settings that you can see here, inpatient care, skilled nursing facilities, outpatient, home health agency, hospice, non-institutional providers, and durable medical equipment. We do not include Medicare Part D.

Our cohort again is aligned with our pneumonia mortality cohort with a couple of additions. We have to exclude admissions for 30 days post --- without 30 days post-admission enrollment in fee-for-service Parts A and B because we simply can't capture 30 days of payment. We have to exclude transfers where federal or a VA hospital is included because we have no method of capturing those claims or payments. And we have to exclude patients who have no DRG during their index admission, and this is vanishingly rare, but because we rely

on the DRG to calculate our inpatient payment, if you don't have a DRG, we can't calculate it.

Number of questions about our transfer attribution methods, so I think it's worth going over. For inpatient transfer patients we define the start date of our episode of care payments as the date of the index admission. So, in this figure you can see that a Medicare fee-for-service beneficiary was admitted with pneumonia to Hospital A on day zero. They were then transferred to Hospital B for the same pneumonia care, and then they made some post-discharge claims and 30 days came up. That whole episode of care payment would be attributable to Hospital A, the hospital that started the index admission.

We calculate payments by removing payment adjustments, we call them stripping or standardizing. Basically, we strip geographic adjustments and policy adjustments, and we standardize where we cannot strip, so we average geographic differences when geographic

adjustments cannot be removed for things like durable medical equipment and other fee schedules.

Next. This is our actual payment calculation example. I don't think I have to run through this. This is the way that Medicare pays for the inpatient prospective payment system. If you move to the next slide, the red Xs show you where we take out the components of geography, IME and DSH. The next slide is an example of how we standardize payment calculations. So this, again, is the January 2012 durable medical equipment prosthetics/orthotics supplies fee schedule. You can see all the HCPCS codes in the leftmost column, so for the first HCPCS codes it's an insulin irrigation syringe. It cost a different amount across four states. We would simply average that amount across all 50 states and apply that average amount any time that claim came up.

We prorate payments, so when

payments begin during the measurement window but end after the measurement window, we only include that proportion of the payment that's included in our measurement window. In this figure you see that a patient was admitted for pneumonia, discharged to a skilled nursing facility, made some claims for home health. Those home health claims straddled the end of our measurement window shown here as day 30. We would assign a per diem claim amount to each day that that person made that home health claim, and only include those per diem amounts that were included in that 30-day window. Next slide.

We do risk-adjust for age and diagnoses present 12 months prior to the admission date and during the index admission as long as they don't represent complications of care. We have a long list that we submitted to NQF. As we all know, we don't adjust for SES, gender, race, ethnicity. These were in compliance with NQF guidance when we developed

the measure. We don't adjust for hospital characteristics or admission source. And we risk-standardize our payments, and this is a numerator of our predicted which is the estimated hospital-specific payment over our denominator which is our expected or average hospital payment for the same case mix that appeared in our numerator. And then we multiply this P/E times the national average for public so we end up in a dollar amount.

I think you can --- if you can advance the slide to the table that shows the same data. Sorry, I wasn't using the microphone at all. These are episode of care payment results. I hope you can all hear me anyway. So, the statistics are in the leftmost column, the unadjusted payment amounts for the entire 30-day episode of pneumonia in the second column, and then our risk-standardized payment appears in the third column. And I just want to point out one thing if you look at the risk in the third column, the median is about $13,000,

but it ranges from a minimum of is it $8,000 and goes to a maximum of about $27,000, so that's a threefold difference for payments made for the same condition after we take away case mix, geography, and other policy adjustments. So, it's a threefold variation in payments we're making for pneumonia care.

And if you can advance to the next slide, I think that's it. I'm happy to take any questions.

CO-CHAIR ASPLIN: Larry.

MEMBER BECKER: I'm sure everybody else knows the answer to this question, and I just --- why did you pick 30 days?

MS. KIM: So, we picked 30 days really when we started to develop these measures we really had an eye toward value, as I think Lein mentioned. We understand payments are one dimension of care, but without payments you can't get to value. Value is really that relationship between payments or cost and quality. And we'd like to get quality metrics

NQF-endorsed, publicly reported, but we can start with any payment metrics, and we really wanted to make sure those aligned, so we picked the outcome that we thought was most inarguable, mortality, can't argue about mortality. And that pneumonia mortality measure is 30 days post admission.

We also thought it made a lot of sense since the attribution is to hospitals and a 30-day window is now an accepted time period in which hospitals are taking responsibility for other outcomes, such as mortality and readmission. It is true that when hospitals discharge patients, a lot of the decisions they make can affect a patient up to 30 days because patients are admitted for conditions that usually require coordination of care after discharge, so we liked it for a lot of reasons. We thought it had a lot of face validity for hospital attribution, and it was aligned with our pneumonia mortality measure.

MEMBER BECKER: So, I didn't hear you

say that it was aligned in any way to the resolution of the condition.

MS. KIM: Correct.

CO-CHAIR ASPLIN: Although, we would hope this condition that might be --- clinically, it would make more sense for this condition than it would with our conversation around heart failure, for example. Good point. Lisa.

CO-CHAIR LATTS: And you may have said this, and I may have missed it. What do you do with cases where the patient is still in the hospital at 30 days?

MS. KIM: We provide them a per diem rate for the index stay. Again, it's not very many patients, but we do provide them with a per diem for their index stay.

CO-CHAIR ASPLIN: Lina.

MEMBER WALKER: Question for the developer. On the call, one of the questions that came up was whether or not --- it was around risk-adjustment, and so one of the questions

was what if there are hospitals that have more ICU beds, and those hospitals are more likely to take in patients the need ICU care? And you said you had some data that you were going to share with us.

MS. KIM: So, thank you so much for the opportunity to clarify. Those ICU analyses were not done with our CMS data. It was done by our group, and it is Safavi in Circulation: Heart Failure. I forgot to bring the reference, but if anybody for Neil is on the phone please note this, and maybe send it out in an email right now.

Just again to clarify for the purposes of this measure, those analyses were done by Yale, but not with Medicare data. They were another project, and it was an example of --- I was using it to illustrate the fact that many times clinical severity does not indicate how providers behave. Certainly, I've had that experience where I admit somebody who has a gastrointestinal bleed, one specialist will do

an endoscopy immediately, one will send home, and one will scope them seven days later, if they have the same exact clinical parameters. So, for me it has a lot of face validity, and that's the reason I brought that up. But I don't want to confuse the Committee. Those analyses were not done on these data.

E.T., I feel like I'm talking to E.T. Yes, hi. Is anybody on the phone? It's Nancy. If anybody has Qian's paper, Safavi in Circulation: Heart Failure 2013, the ICU rates in heart failure patients, if you could send that in to the Committee.

MR. SPIVACK: Hi, this is Steven. We can send that out.

MS. KIM: Thank you.

MEMBER CLARK: Hi. Yes, a question came up earlier, and I don't think you were in the room, but it had to do with the complications that occur during the hospitalization, and you mentioned, of course, that those are not included in the

risk-adjustment. And it sounded like the way you're calculating the payment for those is based on the DRG on the claim, not really regrouping and repricing the claim. Right? So, if that's the case then it would --- if they did have a complication that was not present on admission, it would into a lower paying DRG. Right? Is that the DRG reimbursement you're using?

MS. KIM: So, I'm just going to answer in two parts. We do use the DRG. We don't reprocess the claim or look in any way to parse out the complications of care from the DRG. It is possible. I forget if our materials include the list of DRGs for pneumonia. Your point is a good one, and we should look at that in more detail. We looked at that a little more with AMI because of the procedures bumping up your DRG, so I can't answer that directly, but it's a really good thought, and something that we've been thinking about as we move forward with the other payment measures, on how to deal with

complications that may also impact your DRG.

MEMBER GELZER: Excuse me. Is there a different DRG for complicated pneumonia, versus uncomplicated pneumonia?

MS. KIM: So, DRGS, in general, many DRGs for pneumonia. There are many DRGs for one type of condition, and typically what bumps you into a DRG is your complications, your co-morbidities and procedures. And for the pneumonia ones, I can't recall off the top of my head what they are for the DRGs, but there's more than one. Correct, there is more than one.

And it's not just complications that can bump you into a DRG, it's also if you have any procedures, or if you have a lot of co-morbidities, so it's a little bit difficult to parse out, if we wanted to go that route. But it's a really good question that we have to consider as we move forward.

CO-CHAIR ASPLIN: Very good, thank you. And first a question, Janice, have you joined us on the phone? Are you on mute,

Janice? Okay, so then we will have Jim serve as our lead discussant here as we go through, and then Andrea can make comments from the TEP, where applicable. Jim, do you have --- we move to importance to measure and report. And if you would go ahead with any comments you have.

MEMBER NAESSENS: Okay. Kind of summarizing all the comments and my own perspectives in terms of impact, it's clearly one of the leading causes of hospitalization on Medicare patients, accounting for a lot of costs. Opportunity, there is variation in costs for the same mortality and readmission in the data presented. And in intent, it's intended to provide a cost measure to help assess value, to pair up with CMS quality measures for pneumonia.

CO-CHAIR ASPLIN: Any other comments on importance to measure and report? Hearing none, let's go ahead and move forward with the vote on importance to measure and report. You have to revive your SurveyMonkey here. We have

three sub-criteria and the final overall measure on importance measure report. All right. Let's move forward to reliability, and Jim, once again, if you have comments on reliability.

MEMBER NAESSENS: Specifications did appear precise, well explained construction. As Kim --- as had been mentioned, it was covering all services from admission through 30 days post-admission. Clinical logic appears sound, and look forward to the TEP comments, because there were some concerns that all the TEP questions had not been answered. They are making adjustments for age and co-morbidity.

Condition categories used for co-morbidities rather than the HCCs for Medicare. There's no adjustment for pneumonia severity or type of pneumonia other than perhaps what the DRGs are doing. They are using one randomly selected pneumonia admission per patient per year, excluding appropriate cases,

one-day stays, patients who left against medical advice, those who are in prior or current hospice and bad data, and their standardized pricing. Do you want me to go ahead on testing, or address specifications?

MEMBER BECKER: Can I ask a question?

CO-CHAIR ASPLIN: Sure.

MEMBER BECKER: You say randomly selecting.

MEMBER NAESSENS: Yes.

MEMBER BECKER: What does that mean?

MEMBER NAESSENS: A patient who has more than one pneumonia admission during the year, it's apparently random. There's one of those cases, one of those hospitalizations is randomly selected so that you don't have multiple observations for the same patient. Now, is that specific for a hospital, or if I'm in three different hospitals with pneumonia admissions?

MS. KIM: It's specific for the Medicare beneficiary.

MEMBER NAESSENS: For the beneficiary.

MS. KIM: Yes, for the unique beneficiary.

CO-CHAIR LATTS: So, each beneficiary can only be in the database once?

MS. KIM: Can only be represented in one year, one time.

MEMBER GELZER: So, a frequent flier can't hurt you too --- multiple times.

MS. KIM: Any admissions for pneumonia in the same year, you would get one. And we randomly select, because we're not sure which one is more expensive. We don't want to bias.

MEMBER NAESSENS: And Andrea may want to comment at this time, in terms of the TEP for the specifications.

CO-CHAIR ASPLIN: Great.

MEMBER GELZER: They agreed that the measure population was clinically appropriate. They noted that the ID is based on ICD-10 code

--- ICD-9 codes. They talked about alignment with NQF-endorsed CMS pneumonia morality measures. They felt that the time period for measurement, the inclusions were appropriate. There were no significant issues from TEP with this measure.

CO-CHAIR ASPLIN: Thank you. Jim, do you want to comment on reliability testing? Mary Ann, do you have a question? I'm sorry.

MEMBER CLARK: Yes, I'm not sure this is now or in a minute, but I just have C-- you brought up risk-adjustment, and it's using the CC methodology as opposed to HCC. And I just was a little bit --- I was wondering if you could just comment on that. Because we just saw NCQA use HCC, and now CMS is changing. And I don't know the reason.

MS. KIM: I can comment on that. We do use CCs. We use CCs because we find they're more discrete. The HCCs --- I wasn't present for the discussion. The HCCs are really concerning ambulatory population to predict one-year cost

outcomes, and they do subsume the lower categories; that is to say, if you have cirrhosis, it might subsume hepatitis A, hepatitis B, hepatitis C.

And some of those things, clearly it makes a lot of sense, it just depends on what you're trying to predict. And for us, a GI bleed that's subsumed into something much bigger like cirrhosis, that that makes a big difference on how you treat patients. Are you going to give them blood thinners? What kind of DVT prophylaxis are going to give, et cetera? So, we prefer the CCs for our 30-day inpatient payment and outcome for that reason, because we felt it allowed us to be more specific in terms of the conditions that we could consider for risk-adjustment.

CO-CHAIR ASPLIN: But these are the CCs that then roll into the HCCs?

MS. KIM: Yes, they are.

CO-CHAIR ASPLIN: Okay. Go ahead, Jim, with reliability testing.

MEMBER NAESSENS: So, for testing for reliability they used a split sample comparing half of, my understanding was, 2009 with a developmental half with the validation half. Then also compared it to the full year of 2008, primarily focusing on the inter-class correlation. The method was appropriate, and the scope of testing appeared good. Agreement at the hospital level was 0.825, which is considered to be quite good. And the overall measure was reliable and independent samples.

There were some concerns expressed across the Committee in terms that the data element reliability was not well described.

CO-CHAIR ASPLIN: Any comments from the Committee, or questions from the Committee around reliability in any of the elements that are listed up on the screen? Nancy.

MEMBER GARRETT: I think this is the right place to bring up risk-adjustment, and some of the conversation we had earlier about the Phase 2 measures. And I think the overall

concern that I have is around the risk-adjustment within the hospital episode. And there's this list of exclusions that might possibly associate with complications, so at what point are you over --- you're not adjusting enough, because that's a gray area there, in terms of what's really a complication and what's patient severity, and really does require more resources to treat.

And I worry that this measure is too --- goes too far in not adjusting for those things. So, to the extent that there are differences in the patient populations and to severity across hospitals, that we're not taking that into account. So, that's my big concern about the measure.

MS. KIM: Thanks, Nancy. I know it's always --- it does come up a lot with claims-based measures, and our approach, you're right, is not completely clean. A lot of --- I hope that a lot of this might with ICD-10 and POA. Right now we don't use POA, present on

admission codes, because they're not terribly reliable in the data that we have, but we do use a year, a year --- any transaction you have with a health care system we use a year for risk-adjustment.

So, you're really only considered a complication if you have a secondary diagnosis on your index admission, and it hasn't appeared in the year that you've been there, that you've been seen and you've had some transaction, some claim. So, we do rely on that, and we also do review that list of complications every year.

Our main goal is to be as fair as possible to the hospitals. And when you see the way that our measures are reported, it's really higher than the national average, lower than the national average, it's pretty conservative. You know, we don't say Hospital A next to Hospital B down the street because we are very attuned to your point. We really do try to be fair.

If there is something cleaner,

please share it with us. But in the absence of something cleaner, we're just not sure how to move on that. You know, we do have a clinical panel of experts who does review this year by year, but I understand the comment. I'm just not sure how to correct the measure for that without going too far in the other direction, but I appreciate the comment.

MEMBER GARRETT: I think the opposite would be to control for the DRG. Right? So, that would be going far in the other direction, but those are the two extremes. That would be clean and easy. Right?

MS. KIM: You're totally right, but we are pretty opposed to doing that, because it's --- it would take away the signal we're trying to eliminate. We really are, while being fair and conservative to hospitals, trying to characterize variation in payments that don't provide an equality.

Payment measure doesn't do that, the payment just provides --- the payment

measure just provides the variation of payment, just gives you that one dimension. And then in conjunction with the already publicly reported measures we hope that that's how they'll be assessed, but we don't want to do that, and we don't want to do that because we believe that hospitals are different, and they do act differently, and that has a consequence for patients and care. And we do want to illuminate that without being unfair.

MEMBER GARRETT: In response to your comment about not comparing Hospital A with Hospital B down the street, I thought the intent of the measure was actually to do that comparison for purposes of payments. Isn't that going to be ultimately the use of the measure?

MS. KIM: No, it'll be reported the same way the other ones are, higher than national average, lower than national average. In the hospital-specific reports you'll be able to see how your hospital compares with other hospitals in the state, as well as the nation,

but it won't be oh, Hospital A is more expensive. And remember, the dollars that you see are not real dollars.

They're stripped, standardized dollars that are --- so, I know that in light of your conversation, early conversation about consumers it will be challenging, but that is not the intent, to say Hospital A is more expensive than Hospital B. The way that it will come out is, again, much like our morality and readmission, higher than national average, lower than national average. And we really want to emphasize that higher doesn't necessarily mean bad. We really don't know. What we're doing is providing transparency in the variation of fair prices, apples to apples. We don't know if a higher payment is bad. We need to compare that with quality.

MEMBER GARRETT: Well, that's helpful but isn't that higher than average and lower than average going to be used for payment purposes, as part of value-based purchasing?

Isn't that the ultimate intent?

MS. KIM: That is not the intent. This is --- this measure is intended to be used in inpatient quality reporting. As far as I know, and CMS is here and they can speak for themselves, that is not our intention. I just want to be crystal clear about that, that is not the intention, as far as we know.

MS. HAN: Yes, Nancy is correct. We are --- we don't intend to use it for payment, because just because this is a measure, you need to look at it together with the quality measures. And the way we report it is not like compare one hospital to the other by dollar amount. No, that's not the way we're going to display it.

CO-CHAIR ASPLIN: Jim, did you have a comment, or did it get answered? Okay. There was some question, and Nancy did comment, we understand, you know, at the macro level a split sample, reliability testing. There were some questions about the individual data element

testing. If you could comment on --- respond to those comments?

MS. KIM: Sure. So, the way we --- we rely a lot on CMS's RAC audits for data element reliability testing. As you know, we use claims data so we use all the variables in the claims data that we feel are reliable. There are some things that we don't use because they are not well populated, where we see a large variation across the way it's being coded in certain hospitals. So, some hospitals will code zero percent of the time, some will code it 40 percent of the time, so that's not one we like to use because we just don't feel that's reliable. Have we dug down and done every analysis possible on that particular variable? No, but that's one way we do go through the data elements and try to assess their reliability.

Another thing we do is we look at the frequency year to year. You know, this is a multiple year sample, and we just try to make sure that there is some stability in those

estimates, but other than that we do rely on RAC audits. I'm sure many of you know that CMS does audit their claims in order for payment purposes, nothing to do with a measure. This is just something CMS does in order to prevent fraud, so they are always looking at their variables and claims, and they do rely on that work that they do, for no purpose particularly for this measure, but we do rely on their claims data.

CO-CHAIR ASPLIN: Thank you. Any other comments on reliability, or questions on reliability? Yes?

MEMBER BECKER: Can I ask a question. It's a little bit --- it goes back to this notion of fit for purpose. So, you've made a decision that you're not going to use this measure for payment. Would you at some point in the future change that decision about this measure, or would you come back here first, if you were going to make that change in your use of this measure?

MS. HAN: I don't know there's a process for us to come back. We do take the measure to MAP.

DR. BURSTIN: Right.

MS. HAN: MAP is the ---

MEMBER BECKER: Okay.

MS. HAN: Okay, we do take the measure --- if we propose anything we go to MAP first.

MEMBER BECKER: Okay.

DR. BURSTIN: Just to be clear, at least in our current state, it wouldn't affect the endorsement decision, whether it was for public reporting or payment. It would be just endorsed.

MS. HAN: And I just want to make sure that --- I can't make decision for the future, so I just wanted to let you know there is a MAP that make the decision, make recommendation for CMS. At this point, we want to use this measure for reporting. Okay? And pair it up with our quality measure.

CO-CHAIR ASPLIN: Very good. Not seeing any cards up, are there any comments on the phone? If not, we will move forward to validity. And, Jim, if you could summarize all the comments that was received from a validity standpoint, that would be great. And then we'll vote on both.

MEMBER NAESSENS: So, on --- in terms of specifications, the specifications appeared consistent with the measure intent. Testing, they did a face validity assessment with their technical advisory panel and the consulting health economist. They did data element validity, based on the chart review of the quality measures.

Exclusions, they appeared reasonable and accounted for relatively small proportion of patients. Risk-adjustment, risk-adjustment factors were assessed through bootstrap samples. Their final model is assessed on split samples on R-squared which averaged about a 7 percent explanation. They

looked at over fitting. The worst performance for the other year looked quite reasonable. Standardized Pearson residuals, how many outliers do we have, and it's about 5 percent per sample. And predictive ratios, even the highest 1 percent was only about 17 percent predicted high. The highest decile was about 6 percent high.

There was some concern among comments that the risk-adjustment was inadequate, the low R-squares, and then also the issue that we might need to consider race, we also might need to consider severity as impacted factors.

Identifying significant differences, the current method identifies just a small number of hospitals that are significantly different than the national mean. Comments about being unclear whether the measure discriminated performance across providers, which we've kind of already addressed. And then in disparities, again,

there is concern that the risk-adjustment for race may need to be added, if we're looking at longer term care pending recommendations of the NQF Committee.

CO-CHAIR ASPLIN: Are there any comments from the TEP in terms of validity on this measure?

MEMBER GELZER: The population, again, was felt to be appropriate. They did have a question about assigning cost associated with transfer patients between hospitals, and I wondered, Nancy, if you would comment on that?

MS. KIM: Sure, thanks. Transfer attribution comes up with every measure. So, you saw the way we do it, admitted to A for pneumonia, transferred to B for same pneumonia, claims post-discharge setting attributed to Hospital A.

The only other two ways to --- and it was a very low proportion, 0.3 percent of the population, which has a lot of face validity. Typically, you're not transferring a lot of

pneumonia patients, unlike an AMI or something like that.

The only other ways to do it are really to attribute to Hospital B, start that. We don't like that, because it's a really non-standard measurement period. Everybody else is starting on day zero, and decisions made at Hospital A. And this is an inpatient at Hospital A, this is not from the ER of Hospital A, or an odd-state Hospital A to Hospital B. This is an inpatient overnight stay. Decisions made there can impact the care that they receive at Hospital B, so we feel they do bear some responsibility. And attributing to Hospital A is also consistent with our transfer attribution approach in the pneumonia mortality cohorts.

You could attribute to B but, again, we don't like that because it's non-standard and it misaligns with our pneumonia mortality cohorts and our other payment measures, or you could exclude, and we don't like that because

we hate to lose patients because when you get to small volume hospitals losing a patient also means that you lose hospitals, so we really try to be as inclusive as possible when it's sensible.

So, to us, A made the most sense. And, again, for pneumonia .38 percent of your cohorts being transferred, so in our minds it's not --- it's negligible, but the approach is the same approach we use for other payment and mortality measures.

CO-CHAIR ASPLIN: Thank you. Any comments on validity, any additional comments on validity testing, Jim or Andrea? You covered portions of it already. Nancy.

MEMBER GARRETT: So, I just wanted to make the proposal that we also did for the Phase 2 measures that were like this, that we recommend that the measure be stratified by sociodemographic characteristics that have a conceptual link with the measure, and that we include that as a recommendation, as per this

going forward. I don't know what the process is for that, but that's my proposal.

CO-CHAIR ASPLIN: If we can manage that right now, I guess it's a bit of a usability thing, but since we're in the middle of it, is there a second to that, if we take that as a motion that is parallel to what we talked about with the earlier measure?

MEMBER GELZER: I'll second.

CO-CHAIR ASPLIN: Any discussion on that point, as it relates to this measure? We're presuming we're going to recommend this, obviously. But if we do recommend, then it would be -- all those in favor then of stratifying by socioeconomic status, please raise your hand.

CO-CHAIR LATTS: Do you guys need some background on what we're talking about?

MS. KIM: I'm thinking I don't, but if there are any specific issues, I'm well aware of the ---

CO-CHAIR ASPLIN: Your intuition is right on track, Nancy. Don't mess with this.

This is about ---

MS. KIM: They were specific.

CO-CHAIR ASPLIN: We're outside of our --- we're in our sphere of influence and our sphere of control. Right? And we are making recommendations about how the measure would be used, not making any comments about the specification of this model or anything like that. Mary Ann.

MEMBER CLARK: Yes. So, the only point I would raise is that in the application when they looked at variation by socioeconomic status and race, including race, there was no --- there was little variation. Right? So, I'm just wondering about that.

MS. KIM: That's correct. As part of the NQF application we did look at quartiles of hospitals and a proportion of Medicaid patients, so that's the way we looked at poverty. You can say what you want, but this is the easiest measure of poverty. If you have others that are easy to use in claims data we're

more than open to hearing that.

And where we did that, so hospitals with the lowest quartile of Medicaid patients and the hospitals with the highest quartile really had very little difference in their risk-standardized payments. We did the same thing for race, and this was hospitals with quartiles by proportion of African American patients, and the median risk-standardized payment from the first and last quartile was not that different, and there was a lot of spread in those data, so there wasn't a clear, clear, clear conceptual imperative to adjust for that.

And, again, when we developed these measures the NQF guidance was explicitly not to adjust for SES surveys when we did develop them, but we did do those analyses as part of the NQF process and we didn't see a huge difference.

CO-CHAIR ASPLIN: Thank you. So, since we didn't have that comment and discussion, maybe I'll --- we have a motion and a second, but we can just close the loop on this.

Are there any other comments on this particular issue around stratification? Nancy.

MEMBER GARRETT: Well, just to follow-up on that, maybe a stratification is that that allows us to see what differences there are, and then have done them, so I still think that the --- I would recommend that we C-- that the stratification happen, because that's the only way we'll understand what the disparities are. So, I don't think that that means we shouldn't do it.

CO-CHAIR ASPLIN: Gene, do you have a comment on stratification, or a different issue? Different issue. Lein.

MS. HAN: Just for curiosity, because there are so many ways, or so many recommendations that people give to CMS about stratification, so when you talk about stratification, you're talking about display stratification or --- okay, so stratify by race? That's what you're recommending? I just want to understand.

MEMBER GARRETT: That's a great question. The proposal is actually more in general than that, it's to say that our Committee, because we're in this in-between time when NQF hasn't released the new guidance on risk-adjustment for these factors, so right now you can't actually include it in the model. So, we're saying we recommend stratification of the measure by the sociodemographic factors, where there's a conceptual link with the measure, and where there's empirical evidence for it.

So, we're leaving that up to you to figure out what that is, as the measure developers with the most familiarity with the data that's available and that kind of thing, but we're not going to be that specific about exactly which variables to use at this point.

DR. BURSTIN: But the --- just a quick question. I mean, they're basically showing analysis showing that there isn't an empiric relationship, so I don't know that we

need to formally make that recommendation on this measure, unless you have a conceptual reason you think there's a specific relation to ---

MEMBER GARRETT: Well, that is a good point, but we don't know what other variables are available, what other things to look at, so we're making the recommendation that we think it's important to keep considering. And the stratification over time will show whether those differences start to appear, because it does look like in the literature there's some evidence that there would be a relationship with race, so I think we're saying we don't want to just discard it, that it's still important to consider.

MS. HAN: So, I understand that you'd really like to see we provide analysis stratify, by strata to see whether there is any difference. And I just want to say that that's what we always do, and NQF requires to do that. My --- out of curiosity is that when people say

stratified, we always wonder stratify displaying all this, or Hospital Compare? Is that what people are looking for? So, just need some information, that's it.

MEMBER GARRETT: Yes, that's exactly the idea, is that it's publicly available and that those differences are available to see whether Hospital Compare, or whatever other method there is, so it's not just within your methodology paper that only we see, but it's --- that we can have the conversation about it.

MS. KIM: Thank you. I just wanted to make clear that we did submit some analyses. The analyses that I mentioned are in the NQF application, if that is helpful. The recommendation is to make it publicly available, not that we should submit more, that we should make the analyses that we've already done publicly available. And I just C-- can I ask NQF, is the application a public document? Okay.

MS. HAN: I just want to make --- to

point it out that it's our responsibility, CMS's responsibility to always monitor the disparity care. This is one of our strategy goal.

CO-CHAIR ASPLIN: So, I am not sure we closed the loop here. So, I guess we're kind of in the middle of that. There's been significant conversation here around the desire to stratify by SES in sociodemographic factors, and also some dialogue that maybe it's not necessary for this particular measure. So, I guess I would call the question for the Committee. Those who would favor a recommendation for stratification along sociodemographic variables for reporting the measure raise your hand.

(A show of hands)

MEMBER BECKER: Would you accept a friendly amendment to add the last part, and that is to make it public?

MEMBER GARRETT: Yes, absolutely. That's definitely part of the intent.

CO-CHAIR ASPLIN: All right. So, those in favor of stratifying for this --- recommending that, raise your hand, please.

(A show of hands)

CO-CHAIR ASPLIN: Those who would not recommend stratifying for the pneumonia measure, raise your hand.

(A show of hands)

DR. BURSTIN: Again, this SES report is still being in processed. It's still going through development. Stratification is always something recommended for a measure we think is particularly disparity sensitive. And that's still going to be the case, I think, post hoc, regardless of what happens with this report.

So, I guess I'm a little confused because I think what they showed you in their application is this measure doesn't appear to be very disparity sensitive, so I'm not sure I understand the push to have them openly stratify a measure that doesn't appear to be disparity sensitive other than sort of their

own ongoing monitoring. But I --- it's just confusing to me.

MEMBER BAYEWITZ: That's why I didn't --- I wasn't voting for it. It's not that I think there's anything wrong. I think we should definitely stratify, generally speaking. Just I think it's too soon to start changing things in applications. It's a lot of work to change visual cues on the website. And if there's no evidence that there's any difference, I don't see why we would say anything at this point. Let's wait and then if we see something, maybe that's when we make a recommendation.

CO-CHAIR ASPLIN: Matthew.

MEMBER McHUGH: I think, just to clarify what the level of evidence is, the evidence around socioeconomic status is about the percent of Medicaid. And that's one measure. I wouldn't take it as conclusive.

CO-CHAIR ASPLIN: Lisa.

CO-CHAIR LATTS: I guess my answer to

that, Ariel, would be that if there is no difference that's still worthy of putting out there. So, it's better to put it out there and show there's no difference than not put it out there.

MEMBER BAYEWITZ: And I would say, more broadly speaking, if we said all of the measures that they're showing in Hospital Compare have that kind of stratification, I totally agree, and I would be fine with that kind of recommendation.

But to specify this particular measure, and to then have a requirement or have a proposal that CMS now change their system for that measure, and specifically a measure where we're not actually seeing any difference, to me that would be a lot of work, and would be, honestly, for an end user, a little bit confusing. You know, why split things out if there's nothing really to show?

CO-CHAIR ASPLIN: So, thank you for the conversation. I mean, to me it's sort of a

sensitivity specificity thing. I agree it's too early. We don't know how the whole stratification in SES is really going to play out, so be broad and conclusive rather than really specific at this point, and then we'll see whether that's the right answer for a broad set of measures, or if it's going to be really narrowly selected, specific measures where we want to focus that effort. And it's too early to tell.

All right. So, we have gone in and exited the alley of stratification. We're back on the roadway of validity and validity testing. I think we've heard from TEP and our lead discussant. Are there any other comments on validity from the Committee? I have one, but oh, go ahead. Ariel. Do you have another comment? Your card is up. I'm sorry.

You know, this almost --- this kind of points out just when you think you know these measures, you suddenly realize you really don't know anything about these measures. So,

somewhat sheepishly, I will admit that I had the construction of this along the same logic as the MPSB, the Medicare Spending Per Beneficiary, MSPB measure, and that I thought the 30-day period started with discharge from the acute care admission and ran out for a 30-day period there, like Medicare Spending Per Beneficiary. Now, somebody is probably going to point out that I misunderstood the measure, too.

My assumption is the reason you started it at 30 day, with the onset of the admission, is that that harmonizes with your mortality measure. On a resource-use basis, though, and a comment was made that if you have a longer length of stay eating up those 30 days, and particularly when you think that 85 percent of the national variation on resource-use and the Medicare Spending Per Beneficiary is in the post-acute period, it seems that you're kind of leaving hospitals off the hook that have potentially longer length of stay.

Now, maybe that just doesn't play

out, but in that case I would almost want a resource measure to harmonize more with MSPB than it would to harmonize with mortality.

MS. KIM: It's a great comment. So, the reason did day of admission to 30 days was for all the things we said, but we also think that index care is very expensive and it turns out --- I don't know if you can pull up my slides again or not, but it turns out that for a 30-day episode of care for pneumonia, 61 percent of the payments made in that 30 days goes to index care, and there is variation for the index care in and of itself.

And this differs across conditions, it's much higher for AMI, very similar for heart failure. We do see a lot of variation in post-acute care payments made for pneumonia, much more so than AMI, and we do also find that when they separate hospitals into quintiles of risk-standardized payment, so looking at hospitals who have very inexpensive payments in one side -- and it would be really helpful

if we could pull up my slides -- and the expensive hospitals on the right, that high payment hospitals tend to send more patients to expensive post-acute care settings such as readmission, skilled nursing facilities, and other inpatient care like long-term care hospital, inpatient psych and inpatient rehab more frequently, so more patients, and at a higher dollar amount per patient.

So, we were interested in both the index and the post-acute because I know a lot of measures are focusing on post-acute, but there's a lot of variation in index. And it turns out index is responsible for the majority of episode of care payment costs.

CO-CHAIR ASPLIN: So, the 61 percent you're saying the majority of the variability, or the majority of the cost? I mean, or both? I mean, obviously, it's going to be the majority of the cost.

MS. KIM: Yes, it's the majority of the cost, and we, in developing another suite

of payment measures, were interested in that variation across conditions. That's one reason we also looked at some admission. It differs across conditions. Right? So, for AMI it's 77 percent of your episode payments are index, so that's one reason we looked at it.

We were interested in index and post-acute, and then we were interested what's contributing to the threefold variation we saw in that risk-standardized payment, the total risk-standardized payment? Was it your index care, or was it your post-acute care? It turns out it's a little bit of both for pneumonia, but the post-acute care is --- it is prominent in pneumonia as opposed to other conditions.

CO-CHAIR ASPLIN: Right. So, it sounds like you're making my point.

MS. KIM: Well, I'm making the point that ---

CO-CHAIR ASPLIN: So, why not then have a full post-acute period and a full inpatient index stay and capture all of it like

we do with MSPB?

MS. KIM: Well, the MSPB is 30 days post-discharge, so you lose ---

CO-CHAIR ASPLIN: Right.

MS. KIM: I think that it's important to have a standardized measurement period, so that hospitals are fairly compared over the same amount of time. If you go 30 days post-discharge, on the one hand you're making the argument well, those with longer length of stay have less post-acute care eligibility, or something like that.

But if you go 30 days post-discharge you also don't have a standard window, so somebody who discharges a patient within two days you're just not --- I don't think it's fair. You don't have a standard measurement period.

Either way, it's a reasonable measure. Obviously, MSPB is endorsed, but if the question is why didn't we go that way, it's for all the things that we said, 30 days

post-admission seems much more attributable to the hospital. Once you go 30 days post-discharge you're getting further from that index admission. We did want to harmonize with our mortality. We wanted a mortality cohort, and we really felt it was a standard start date.

You know, everything that happens in the index also influences what happens in the post-acute care, so if you take that away or you minimize that by extending the post-acute care, it's just another trade off. I'm not saying one is better than the other, but if the question to me is why did we do that, those are the reasons we did that.

CO-CHAIR ASPLIN: Thank you. Jim.

MEMBER NAESSENS: Actually, MSPB covers from three days prior to admission through 30 days after discharge, so you do have that acute piece, plus you have the admission piece.

MS. KIM: That's basically the way a

Medicare --- oh, I'm sorry.

CO-CHAIR ASPLIN: No, go ahead. I mean, I would just --- to channel Jack again, I mean, I --- his comments on both of these earlier, and on MSPB around how much the DRG consumes the variability for the index stay, you know, but you're saying that there's not as much --- there is quite a bit more variability even in the context that you ---

MS. KIM: MSPB is all condition, so again the DRG is going to weigh them a little more heavily. We're condition-specific, so everybody in ours has pneumonia, so there is still play in the index variation. There will be less because of the way the two measures are constructed.

And just to comment on the three days pre, that's the way Medicare pays through the IPPS. We didn't do that because related payments now they're considering one more related three days within your patient stay is kind of bundled into the way Medicare pays for

your inpatient stay. But some of that stuff isn't really related, like an ambulance cost or something, so we again wanted to be fair to the hospitals and not make their payments look more expensive because stuff that happens before your admission, it's hard to hold the hospitals liable for that. If you decide to take an ambulance to the hospital, that's expensive, and we didn't want to jack up the costs, forgive the colloquialism, to the hospital, because the hospital didn't decide to do that. So, that's the reason we don't do that.

CO-CHAIR ASPLIN: Very good. Are there other comments or questions from the Committee relative to validity? Gene.

MEMBER NELSON: I was just wondering how the propensity to admit patients for pneumonia, and how that varies across hospitals might play into the --- how to interpret the results for cost? And the same would apply to the other paired conditions, or paired measures.

MS. KIM: I think that's a great question. I think the propensity in it probably plays a lot into it, you know. What we know from other literature, again not directly related to this measure, is hospitals do have phenotypes, they have propensity to admit, propensity to readmit, et cetera, et cetera, that may or may not have anything to do with the clinical severity of the patient.

And maybe, again, we haven't investigated this with this measure, that may explain some of the variation we're seeing. So, we don't know to directly answer your question, we're not sure, but I --- my hypothesis, untested hypothesis, is that it may explain some of the variation we're seeing.

MEMBER NELSON: This may not be correct, but you would think that if your propensity to admit is higher, then severity is going to be lower, costs would be lower, bigger denominator, better results. High cost, better results and quality.

MS. KIM: Yes, we're concerned about that. Again, now we're just in the realm of hypothesizing, if you're asking what I think. It is true that a lot of those low severity cases for pneumonia, those may be one-night stays or going to Obs, so those Obs --- the way Medicare treats Obs is like an outpatient, so those don't get into our measure.

You're only in our measure if you're admitted as an inpatient for pneumonia. If your hospital considers you an observation, and many pneumonia patients are observation stays, that doesn't get you into a measure. You have to be an inpatient, so it may be taking care of itself, unless they have a propensity to stay. Admit and stay, and we haven't done those analyses. It's hard to drill down in claims.

CO-CHAIR ASPLIN: One of the few silver linings of two midnights. Larry.

MEMBER BECKER: So, the deeper we get into this the more confused I get. And I'm sure this is naive, but my assumption about this

whole 30-day readmission piece was that we want to do the right things in the hospital to get the patient to a place where the probability of them being readmitted is as low as we can reasonably make it.

And if that's the case, then we're counting this period of what it takes to get somebody to that point, irrespective of their --- you know, because all these different severities. So, why does the clock start back here? Why doesn't the clock start once I get you to a place?

MS. KIM: Because our conceptual model is that of an episode of care payment, obviously, re-admissions make you more expensive, but it isn't really predicated on trying to harmonize with our readmission measure.

That's a different way we could have gone, but we chose mortality because we felt it was, quote-unquote, the harder outcome. And the mortality measure, the pneumonia mortality

measure is day of admission plus 30 days. So, just to clarify this is really a payment measure.

I know we've been talking a lot --- our stream of conversation has flown a little bit, but this is really an episode of care payment measure that's harmonized with our mortality measure. And that's the reason for day of admission to 30 days. That doesn't directly answer your question. It's another way we could have gone but we didn't because, again, we really had to make the choice.

The reason we didn't choose readmission is it conflates the payment. Re-admissions are expensive pieces of your episode of care payment, and it's sort of tautological. And we really wanted to minimize that tautology in this first go around and have something sensible that's harmonized with a hard outcome, like mortality.

CO-CHAIR ASPLIN: All right. Many choices along the way, great. I think we're

going to --- unless there are other comments or questions. I'm not seeing any currently. We're going to move ahead to voting. You have your survey in front of you, and ask that you would make your recommendation on both reliability and validity, keeping in mind the conversations we had along the algorithms this morning. Okay. Jim, could you take us into the next section around feasibility, please? And your comments and the comments from the Committee.

MEMBER NAESSENS: And on feasibility, basically, it's based on Medicare claims data, so all the burden is on CMS.

CO-CHAIR ASPLIN: Good. And, Andrea, any comments from the TEP relative to feasibility?

MEMBER GELZER: No.

CO-CHAIR ASPLIN: Any comments from the Committee relative to feasibility, any questions? Seeing none, let's go ahead and vote on feasibility. And, Jim, you can summarize

comments on usability and use, please.

MEMBER NAESSENS: For usability, generally, they said it would be useful for providers. Kind of the challenges or the questions were that it's not really useful for consumers. It would be ideal if they could hold the hospital and post-acute care provider jointly accountable.

This will be a challenge for providers as transparency as possible in the utilization outside the index provider must be accessible and easy to use, including specific reimbursed amounts and providers by name. In other words, we can't address the outside providers unless we know where the money is going and who they're going to.

CO-CHAIR ASPLIN: And any comments on usability from the TEP or Committee? Nancy.

MEMBER GARRETT: So, just a question, actually, about that comment. I think that's really important in order to make this useful, to give the providers that really

detailed data, including costs. Can you tell us a bit about the plans? I think for the last measure you showed us some of the information that providers will get, and I can't remember what it had in it.

MS. KIM: Thank you for the question. So, the last time I brought to the Steering Committee an example of our hospital-specific report, this was particularly for AMI, but it will look very, very similar for pneumonia. And it should really be a reflection of, also, what you get for mortality, but it will have more complicated cost data.

It will have pages, it's an Excel spreadsheet, and each hospital will get a --- the first column is your Medicare beneficiary that fell into the measure, so all your pneumonias for the period of time that the measure, this is '08-09, so you get all your pneumonia patients in your measure for 2000-2009, the first sheet will be a summary. You total up the payment how it breaks down by

the index payments, as well as post-acute payments. And the next tab will be all the post-acute care, more information than you ever wanted or could make sense of, but you will be getting a lot of specific information on who went to what setting, and how much was spent in that setting.

I should have brought another page, but if you remember from last time for those of the Steering Committee, it was very detailed, line by line of every claim that that particular patient for, in this case pneumonia, had made in that episode of care, whether they died, whether they lived, where they went, what they spent there, and whether they were readmitted, as well as if they had Obs stays, ED stays.

We really try to make it usable for the hospitals, so we try to categorize the claims in usable buckets for the hospitals.

MEMBER GARRETT: So, if you could convince the people doing the Medicare Spending Per Beneficiary to do that, too, that would be

great, because they're not giving us --- we can see what other providers people are going to. It's just the top 10 per service category, and it doesn't show you the money, so you're really kind of at a loss of how to take action because the data is too general.

MS. KIM: I can't comment on that a lot. I'm not involved in MSPB at all. But, you know, again when you see the files I hope that it's usable. We tried to really fashion it from the hospital's perspective as the hospital's. These are questions that I would like to know if my patient was readmitted, if they went and had an ED visit, an Obs visit, an outpatient visit, whether they went to a SNF or rehab. So, it's really focused on those types of post-acute care settings, meaningful. I think they're clinically meaningful.

CO-CHAIR ASPLIN: Gene, and then Larry.

MEMBER NELSON: I have --- I was looking at the sort of the three by three table,

high dollars average, low dollars as the columns, and the rows are high quality, medium quality, low quality, and we fill it in with 5,000 hospitals. And how many are going to fall in some of those interesting quadrants, and how useful would it be if the number is vanishingly small, that are, for example, high quality-low cost, or low quality-high cost?

MS. KIM: So, to answer your question, we don't know yet, we've bootstrapped those results again to make sure that we have an estimate with a certain --- a range of uncertainty and confidence in the formula that before we categorize any hospital we have to do that.

We feel we have to do that, again. But for AMI just to give you a sense, I think I can share those publicly. Right? I shared them with the Committee last time. It was: 8 percent were higher than average payment, and I think 3 or 4 percent were lower than average payment, so that's the magnitude it was for AMI.

We won't know yet for pneumonia, but I will say for the risk-standardized payment in AMI we saw a twofold difference in risk-standardized payment from the min to the maximum, and then we saw threefold, like it went from $8,000 to $27,000. So, if that's what it was for AMI, I would imagine it would be more hospitals in pneumonia, but I don't know for sure, to answer your question directly.

MEMBER NELSON: But once you pair that with the quality outcome it gets very --- that 8 percent gets really spread out.

MS. KIM: Just to be clear, there's no statistical method to incorporate both the morality and a payment outcome right now. You would have to look at --- you could go on Hospital Compare, you'd see their payments for each hospital in quality, mortality and readmission estimates.

And it would be more of a qualitative view, so we don't provide those quadrants or anything on Hospital Compare. It's

confusing and difficult to interpret, so you won't see something like the quadrant there on Hospital Compare.

CO-CHAIR ASPLIN: Larry.

MEMBER BECKER: So, twice you used the years 2008 over 2009, or vice versa. I assume that was for testing. So, what's the time delay when hospitals actually get their data?

MS. HAN: To answer that question, for the claim-based measure we usually have one year lag, nine months.

MS. KIM: It'll be 2010 to 2013.

MS. HAN: Yes.

MS. KIM: But it's usually one year lag.

MEMBER BECKER: Okay.

CO-CHAIR ASPLIN: Lein, did you have a comment? Did you have other comments to make?

MS. HAN: Okay. I don't know which question right now, because it's kind of delay now. So, I think the question was related to the display quality and also the payment. I think

this is not an easy thing to do, so we talk about value of care. How do we splay value of care? So, this is something that CMS is working on. We do consumer testing and try to convey the concept.

Welcome recommendations. I mean, if you guys have any recommendation to splay this. We also try to do some environmental scan to see how other organizations are currently doing that. It's just that's a new thing, and then it's not an easy way to display this.

CO-CHAIR ASPLIN: Any other comments? Gene or Nancy, do you have any other comments? Nancy.

MEMBER GARRETT: Just a quick thing on recommendation. It's a recommendation for a little bit of different thing, but again talking about the data that you're giving to providers to help make this actionable. One of the big challenges, as a provider, is that we're getting data from CMS, as well as from other payers in all kinds of different formats, and

so to have the resources to take that data in and make it usable, it's really unaffordable.

So, to the extent that, like, within CMS you could harmonize the data that you're giving to us for the different efficiency measures, so Medicare Spending Per Beneficiary, these resource measures, that would be a huge step forward, or even creating some kind of interactive website where you've already created a queryable format for us, so that we're not having to invest in those analytic resources to set it up so that we can start looking at it. So, that's just something that would be really helpful.

MS. HAN: Yes, this is a great suggestion and I appreciate that you point it out. I just want to explain one thing. I'm not familiar with that measure, but I want to say that whether we have the authority to provide what kind of data, especially patients from other hospitals, other care settings, it all depends on the program, whether the program

that we have the authority to do that.

For example, for IQR, HRRP we have the authority, we can do that. But others that we need to check into that, so that's probably sometimes you got limited data for certain programs, but you got a lot from HRRP or from IQR. That's what happened. And whether we can combine together, you know, we just have to check legally whether we can do that, because of HIPAA, you know, all this concern. I just wanted to make that clear.

CO-CHAIR ASPLIN: Unless there are other questions or comments from the Committee then I would call the question, and ask you to vote using the survey tool on usability and use, as well as your overall recommendation for endorsement of Measure 2579. And on behalf of the Committee, I'd like to thank you both, Nancy and Lein, for your comments and being here to answer our questions. Really appreciate the work that's gone into the measure.

MS. HAN: It's wonderful discussion,

and like to see, and great that you are engaged in this. This is wonderful, thank you.

MS. KIM: Thank you for the opportunity.

CO-CHAIR ASPLIN: We always have to find something new, Nancy, to dive into. A little less on the risk model this time, but we got into the 30-day post-acute period.

CO-CHAIR LATTS: We wouldn't want you to be bored.

MS. KIM: No, I wouldn't want to be.

CO-CHAIR LATTS: So next time be prepared to defend that.

MS. HAN: We have a very good contractor too, they did fantastic jobs.

CO-CHAIR LATTS: Great, thank you. So, guys, we're done early, obviously. Now, the options are, a), continue pressing on until 5:00 doing some of the work tomorrow and ending early tomorrow, or leave now and go later tomorrow. And my vote is to end early tomorrow, and I'm seeing nodding around the heads, so with

that in mind let's take just a five-minute break, and then we'll resume and go for another hour, and then we'll plan on ending at noon tomorrow, with lunch for folks who want to stay and have lunch. And for those of you who are fortunate to live locally, you can get out of Dodge and go do stuff.

(Whereupon, the above-entitled matter went off the record at 3:57 p.m. and resumed at 4:13 p.m.)

CO-CHAIR ASPLIN: Well, recognizing that we've probably lost our mojo for the day, if we don't get started again here quickly we might as well just pack it up and head to the hotel. And given the expense of bringing us all here, we do want to make sure we devote sufficient time to the dialogue we had dedicated for tomorrow, so to honor that we had the conversation earlier this afternoon, and we can --- I guess I would ask Ashlie or Lindsey to help us frame up exactly what the most important next steps are in this overall

conversation around direction setting for cost and resource-use measures.

MS. WILBON: Sure. That was a hand-off to me. Correct? Sorry, I was ---

CO-CHAIR ASPLIN: Yes, that would be great.

MS. WILBON: Okay.

CO-CHAIR ASPLIN: I'm thinking as I'm speaking here. We ended up on page 4 of the discussion guide having a number of different factors. Anyway you could display your ---

MS. WILBON: Okay, sure. So, just based on where we ended up over kind of keeping with the discussion guide. And I want to just note, Taroon has joined us for this discussion, so he'll be chiming in as well. But we --- trying to be flexible with today's schedule we looked at what we have left in the discussion guide, and we would like to keep the latter part of the discussion around linking costs and quality measures, and the criteria piece to tomorrow just because those two pieces kind of go

together. Andy Ryan will be here tomorrow. He's one of the authors of the paper around linking costs and quality, and he's going to help us through that discussion, so we really want to make sure he's present for that. So, we're going to save that piece for tomorrow, and that's probably the more lengthy part of the discussion. But we've broken off this piece right here that we think will probably take us between now and 5, or we may finish early. But as sort of a primer for tomorrow's discussion in terms of the path that we're on, and whether or not we're still --- our goal is still the same in terms of getting to efficiency and value.

We have this conceptual model here that we've been using for some time that I'm sure everyone has seen a million times by now, but we wanted to just take a step back again, look at this conceptual model, and see whether it still fits where we think we're going in this space in terms of using resource-use measures

as a building block to get to efficiency and value. Is that still our goal? And just to make sure that, you know, are there any pieces of this model that we need to add to. We've somewhat added a dimension to the work that we've been doing here at NQF in terms of including affordability and what that means from a consumer perspective. As staff, we've been exploring ways and how to integrate that affordability concept with this conceptual model, so ideas that you have on that are also welcome. So, really just wanted to get Committee reactions on the path forward, are we still on the path to efficiency and value, and any input in particular on the conceptual model as it relates to that in terms of the path forward. And I think that will set us up nicely for the discussion tomorrow around actually linking cost and quality measures, assuming that we are still on the path to efficiency in terms of those --- combining those two signals. So, I will leave it there and ask if Taroon has

any additions or clarifications.

MR. AMIN: Yes, I think the only thing that I would also like to raise for the Committee's discussion is that, you know, as part of this conceptual model that we laid out a number of years ago, we looked at sort of resource-use with the idea that we would be endorsing resource-use measures in the context of quality, and with the implication that efficiency could be evaluated in an objective way. And one of the things that we've asked Andy to do and, again, we'll have some version of this conversation tomorrow when Andy is here, but I know, Cheryl, you on the phone, as well, have done a lot of work in terms of profiling efficiency models that have been used across different types of applications. But, you know, the idea here in terms of the way we've framed efficiency is that there should be some objective way to really look at quality performance in the context of cost performance.

And one of the questions that the

staff have been thinking through or conceptually trying to rationalize is whether such an objective evaluation is actually possible. When you're looking at sort of quality and cost constructs, is it really possible to objectively put them in a single model to be able to come up with a single score, or to be able to display them in a way that comes up with some summary score for the purposes of a payment or even a public reporting program, or is just the fact that you're putting these two constructs together imply some type of weighting. What quality measures you select, how much you weigh the quality measures vis-a-vis the cost measures. There is some implicit weighting of these two different constructs, and that weighting would seem to imply that --- would seemingly vary depending on the stakeholder.

So, one of the specific elements in addition to what Ashlie described, which is essentially our journey, you know, the journey

that we've described in the past and we've talked about a number of times is that, you know, the importance of cost and resource-use measures, while they don't offer directionality in themselves, necessarily, of which way is better, we use the quality measures to give us the fuller picture. And we're on a journey toward efficiency and value.

But I guess the question really that we're trying to explore is, you know, how --- is there this objective weigh station, if you will, of efficiency before we get to value. And is that really where we're trying to get toward, or once we have these cost measures, and once we have quality measures, you know, what is the ideal state for how these come together? And is it at the measure level, or is this something really that ends up being at the program level of how one would put these two concepts together to be able to give reliable and valid estimates of cost and quality performance together.

So, again, that's a lot to start the

conversation, but that's the nature of the conversation we want to get toward just to make sure that we're all clear on, you know, where our end destination is, and making sure we're, you know, making incremental progress toward getting there.

CO-CHAIR ASPLIN: All right. Lina, you want to start?

MEMBER WALKER: Sure, thank you. Just reflecting back on the more recent measures we've evaluated, the way that it's been presented has been around these categories of --- you know, around average, above average, less than average. And the average category is very large, as we just heard for the AMI measure, you know, 90 percent of hospitals fall into that category. So, inherently we're looking at measures, current cost and resource-use measures and quality measures that are very imprecisely measured. You know, there are these broad categories where most of

the hospitals for those particular measures fall. So, if you're using those kinds of constructs, I think it would be very difficult to develop an index measure that would represent efficiency. Now, that's --- you know, so if this is the direction we keep going in terms of evaluate how we categorize and evaluate cost and resource-use and quality, I don't think if we wanted to get to an objective measure or an objective index, which I personally don't think is the direction we should be going, we're not heading in that direction right now.

CO-CHAIR ASPLIN: Thank you. Gene.

MEMBER NAESSENS: I guess first a disclosure. We've written a book, Value By Design, and so I'm not plugging it. But it is to say I think I've thought quite a bit about it, and probably have my biases that I'll share now; that the idea, as Taroon said, of in health care and looking at the product of health care as being health outcomes, having a technical

one number measure of efficiency, to my sensibilities, is not the right way to go. That, again, if we're looking at from a community or social perspective, what are we trying to get better health outcomes, lowest total real cost to patients, families, communities, society. So it's end use best health results, best outcomes at total lowest cost possible.

So, a community might make decisions about better outcomes and lower or acceptable costs, but for an individual, I've got cancer now, it becomes highly specific, and highly vested interest, and society might step back, and the economist might step back and have guidelines about cost per quality- adjusted life-year, et cetera. So, I think an approach is to have our best indicators of health outcomes especially that matter most to people, and alongside those, literally alongside those have our best estimates of, as I mentioned earlier, total direct cost and total indirect

cost. And then enable an individual to make their decisions, and for health systems to try to literally, or provider groups literally, or health plans literally to try to find best ways of getting the best outcomes at the lowest production costs and, therefore, pass those lower production costs over to the people that are paying for it, the family, the insurance company.

So, that is to say a value array that has health outcomes, oftentimes technical indicators of quality and especially harm, and cost indicators to patients and the community that can be looked at side by side is most helpful, rather than doing let's say, a cost per quality-adjusted life-year approach.

As I was talking about earlier for spine surgery, we can do that, and we have done it, show the incremental cost for per quality-adjusted life-year for surgery, and so we can come up with one number. But it's probably not all that helpful to all that many

people unless we're going to ration and make our list and say we're only go down this far in our list of interventions at x cost quality-adjusted life-year. So, a value array rather than a number, especially focusing on outcomes and cost.

CO-CHAIR ASPLIN: Nancy.

MEMBER GARRETT: So, I really agree with you, Gene. I like the way you eloquently stated that, and I kind of think about consumer reports and the way they present ratings. And so you get information essentially on quality, let's say you're buying a dishwasher, you can see quality arranged from high to low, and however they're --- whatever dimensions they're including, but then you also have costs. So, you might be looking for those low-cost high-quality items, you might be in a situation where you actually want to buy a brand name so you want to go to Mayo, and that's what's important to you. But it gives those decisions to the consumer to make, so I think having those

dimensions available in an easily consumable format is really important.

At the same time, I think we're in a situation because of the rise in health care costs and the need to do something about it as a society where increasingly employers, health plans are going to be making those decisions for consumers. So, narrow networks, for example, who's going to be in the network? Well, it's going to be those low-cost high-quality providers that are going to make it in the network, and so that's going to influence the choices available to a lot of consumers. So, you've got both factors going on at the same time, I think.

CO-CHAIR ASPLIN: Carolyn.

MEMBER PARE: I think what you say, Nancy, is right on, and employers and health plans are making those decisions for consumers right now, but there's a lot of suspicion around that as to whether or not there are any quality factors in those kind of limited networks. So,

I think it behooves us to be on a fast track to array the quality metrics, as well, so that we can dissolve that suspicion, as well as start making the point that you can get high quality at low cost.

CO-CHAIR ASPLIN: Backing up to the broader framework question here, I really agree with the comments that have been made. One comment I would make is that it really just depends on who your audience is as to whether the framework, because it's not the same for the various constituencies. I mean, I could create a story for a delivery system that wants to take on risk for a population that would look different than the framework that the very populations that organization is trying to serve might think about these issues or the plans, et cetera. So, it depends on your constituent group.

I would say that I agree with Gene that the best combination of quality, experience, engagement, perhaps health

behavior, and functional status measures on one side coupled with as specific as possible indicators on cost over time or cost related to a bundle is really going to be the most helpful story for a period of time. And down the road, you know, I know I've advocated for this even earlier today, at some point we may get to meaningful descriptors and measures of what it costs us to create healthy days, et cetera. Right now the ratio isn't as important to Gene's point as having compelling cost, quality experience, functional status measures alongside --- I think I meant quality, I said cost, alongside specific cost measures. So, a comment on the latter group. It has to go a lot further than average, you know, above average, below average where it's kind of a Lake Woebegone story. Just only that everybody is average, not above average. We've got to get more specific if it's going to be meaningful, meaningful discriminators on cost, I think. Lina.

MEMBER WALKER: One of the challenges of getting more specific with the cost measures is precisely some of the challenges we've raised in this Committee. Like if you don't have the right data to risk-adjust appropriately, then you have to allow for those errors. So, in the cases that we evaluated in heart failure, AMI, pneumonia, it made sense that they had these wide ranges because you acknowledge that your risk adjustment is imperfect, and so you allow for variation and let them all fall in the same group. I mean, I think it's a high bar to say we need to get more specific. You know, until we get the data, I don't know how we're going to be able to achieve that bar.

CO-CHAIR ASPLIN: Wouldn't some of that have to be, you know, marching to --- with the eMeasures and so forth using more clinical data, and that's been commented on during our last meeting that we were face-to-face. We won't get there with claims only cost measure.

Yes, Larry.

MEMBER BECKER: So, I was particularly bothered by, I don't know what she said, 2 percent above, 6 percent above, 4 percent below, everybody sort of hiding under the yellow line. Doesn't that defy the laws of statistics, that there is an equal distribution of performance? And isn't the point to give people information so they can improve their performance? If we allow everybody just to hide under the yellow line in the middle of the road, nobody is incented to make change and to get better.

CO-CHAIR LATTS: It also suggests that their stratification has removed all variability when they risk-stratify all those various pieces.

MEMBER BECKER: Right.

CO-CHAIR ASPLIN: A lot of that depends on what happens with the business model. Right? And the measures, the measure framework is one component that will push the

business model. I think the market forces will take care of other aspects of it. Lisa, and then Ariel.

CO-CHAIR LATTS: So, one of the things that sort of bothers me is we're talking about the sort of conceptual framework here versus the real world applications. In the real world, cost and quality tend to be a floor, so especially for cost, you pick your measures. You have your basic standards, and you're not, at least in my experience and those of you working with health plans can contradict me, but you're not necessarily picking your high-quality providers. You're picking the providers that are above a floor and then looking at cost. So, it's not that you're picking, necessarily, your high-quality providers, you're making sure that the providers that are a) willing to partner with you because that's number one differentiator, b) coming in at a reasonable cost mark. Their quality is not substandard, so it's a floor

versus a ceiling.

CO-CHAIR ASPLIN: Ariel.

MEMBER BAYEWITZ: Yes, so just --- I'll move back on that last point. So, I think it probably depends on the use. I will say from a plan perspective it is not just about a floor, it's a ladder, and we do have significant variance in terms of how we pay providers and how we share risk, based on how you land in that ladder. And it's not if you're average you get this, and then we look at cost. It's if you meet the floor, then we look at you period. We won't give you anything if you don't hit the floor. Beyond the floor we'll give you more the better you get.

Now, not all programs are like that, but I would say, I mean, most of the ones we have now are like that. Just in terms of, Larry, your comment, you know, I have to say for a lot of the measures that we looked at, it sounds like what she was describing in most of the markets, especially when you talk about quality. I mean,

most providers do hover around that middle point, and then you do see people on the tails. Oftentimes the tails are the smaller groups, and you kind of wonder is it because they're really bad, or is it just because they're small size. But generally speaking, the large groups do tend to cluster, it does vary by market. And I wouldn't be surprised on the resource-use if it was the same. It would be nice if you could have a wide distribution, but a lot of the measures that I look at it's not like that.

MEMBER BECKER: Are the big groups clustered because they're big groups and you can't discern the performance inside the group, because when you take 100 docs, you got a normal distribution, and so they cluster around a mean?

MEMBER BAYEWITZ: It might be. I'm just saying from a group perspective, I mean, it's --- and it's not just the really large groups, the majority of the groups are, you know, tend to go towards that. It could be what

you're saying, it could be that because you have a wide distribution within an organization and people practice differently. It just --- you know, the smaller you get the harder it is going to be to evaluate them.

CO-CHAIR ASPLIN: Carolyn.

MEMBER PARE: In our experience it is. There's wide, huge variation in those big medical groups, and so you do have --- when you spread them all out, you do actually have small practices that are doing extremely well in comparison to bigger groups. There's tons of variability within those medical groups.

CO-CHAIR ASPLIN: Yes, there is. And the same phenomenon occurs within the group, though. I mean, we found that our overall D5 compliance score got higher the larger the panel size, in general, with some exceptions. So, same phenomenon occurs within.

Other comments or questions on this aspect? Are we answering the --- yes, go ahead, Cheryl.

MEMBER DAMBERG: This is Cheryl. I just wanted to jump in and say that, you know, I always like a conceptual framework, and I think it's useful to have this here to remind us what we're trying to achieve. But I feel like we're having a difficult time just getting out of the green box, let alone to work our way up the scale. And I think part of that is, you know, we've got sort of more than a decade, almost two decades worth of a lot of foundational work around measuring quality. I think we are still in the very early stages of just even trying to figure out how to measure resource-use, and what it means, and how to display it, and how to use it. So, I think in order for us to be able to move up the scale there's going to have to be a lot of what I call conceptual methods work done to really advance the space. But I guess I would ask NQF and maybe the Committee to sort of take two steps back and say, you know, what are our near-term goals? And is our focus here really on trying to help reach the three aims

of the National Quality Strategy, one of which I believe is reducing spending. Because I think that a lot of the measures that we've considered start to get at that by putting providers on notice that we are going to be looking at their relative resources spent, you know, for a given year, or condition, and I think that starts to move that discussion.

I think we are probably a little bit premature to put it side by side, but I think we're getting a little closer to that, particularly where the measures line up well with whatever the resource-use measure is.

CO-CHAIR ASPLIN: Thank you. Gene.

MEMBER NELSON: It's probably because Jack Lundberg is down the hall from me, but I would take the opposite point of view that was --- what was just said, that we've had since the late '70s, looking at hospital service areas across the country, total Medicare expenditures per person per year based on what hospital service area you're in, and we see

fivefold variation for many conditions, but there are some conditions that are very low variation, and many that are extraordinarily high. And getting that actually per person rolled up annual measure is extremely helpful, and then with that, and that being part of a reporting and payment environment, if you're Dartmouth-Hitchcock Health System and you start getting measured and paid for better outcomes at lower per capita cost. We're free to innovate, so we can do things because what used to be a profit center is now a cost center. So, to have these things that we've been studying academically with the Dartmouth Atlas and others have done around the world, it's very helpful. And we know a lot about these higher level public expenditures that we might call public costs, like Medicare, spending per person per year, case mix adjusted. So, I take the opposite point of view that we actually ---

MEMBER DAMBERG: Gene, I'm sorry. Actually, I wasn't trying to be inconsistent

with what you just said. I agree that we know a fair amount about differential spending and variation, and I would like to say that we suggest the price you're talking about which is better outcomes on a lower per capita cost basis, but I think we're still struggling with how to measure outcomes.

Now, maybe within health systems that's sort of an easier task, but I don't feel like we are quite there yet. I think it's a laudable goal and something we should be working toward. And I guess the question is, is does the conceptual framework on this slide that's being displayed capture that?

CO-CHAIR ASPLIN: Thank you. That's a good clarification. I think the presumed bucket that would capture that would be in the generic term quality. Right? And I think what you're suggesting is that it needs to be more robust than what we have today. Joe.

MEMBER STEPHANSKY: Just a couple of things that are unrelated entirely to anything

we've been talking about. As I have already said, the consistency, the value of consistency is highly overrated.

In terms of looking at that conceptual box, and this goes right to something that Gene was talking about earlier, and that we talked about in the Linking Quality and Cost Committee, was that it's a matter of whose costs are going to count in that, because right here we're talking about the cost resources used to provide care. But when you start --- and not all of them are in there. And we have to make choices about if we have efficiency as a ratio of outputs to inputs. Well, who gets to decide what counts as an input? Are we going to count family costs in this? Are we going to count transportation cost to consume care? So, that's a whole different area that perhaps that resource-use box needs to be a little bit more fluid.

The other thing that we got to talking about in the Linking Cost and Quality

Committee was there is some evidence, and perhaps a growing body of evidence that we cannot keep talking about higher quality is going to happen at lower cost. Higher quality may mean higher costs, and we're just shifting around where the dollars get spent so to speak. I just want to be careful about that, that high quality-lower cost kind of concept, and that's not universally accepted.

CO-CHAIR LATTS: Well, I think there's --- just to comment on Joe's comment. I totally agree with you. I think that the whole paradigm that high quality will lead to lower cost is --- I'm not a believer. But I do think that it's then high quality along the range of costs, and so you're able to select the lower cost providers within that high quality bucket, theoretically if they exist.

MEMBER STEPHANSKY: Yes, you were assuming that somehow that that quadrant, the low cost-high quality quadrant has some providers in it. And there might not be anybody

in it.

CO-CHAIR LATTS: Well, but I think the question then gets to be if you build it, will they come? So, if you incentivize correctly, you know, whether it's through reference pricing, or some sort of differential payment that if you build it, will the prices drop appropriately such that there will be someone in that quadrant? And I agree with you, it may be, you know, the Yeti phenomenon.

CO-CHAIR ASPLIN: Mary Ann.

MEMBER CLARK: So, in terms of this graph, again back to that, I mean, we have this big efficiency bucket, as well. And I guess when I look at that I think of tying it into costs. And you mentioned well, maybe there's nobody in the low cost-high quality bucket but, you know, when I think of efficiency, I almost think of it more from a internal facility efficiency and process improvement because, you know, we heard from --- even from CMS that the highest cost of these episodes is the inpatient episode. And

hospitals don't really know what it costs them, you know, to provide their services, and we're basing costs on reimbursement right now.

You know, I think there needs to be somehow, maybe we need to talk about more emphasis on doing more, you know, the time-driven activity-based costing studies or something in order to get at well, how are these hospitals actually going to manage their costs actually, you know, drive it down to the payers, to the employers?

CO-CHAIR ASPLIN: Larry.

MEMBER BECKER: So, with all due respect, you might be right eventually, but I think there's so much variation in the system now that there's got to be efficiency, and we see it in the cost distributions, you know, in the stuff that Lansky did on the West Coast with colonoscopies, as an example. So, there's always variation in the system, and until we get to an actually efficient system --- and when we do, I think you're probably right. But in the

meantime, there's all this waste in the system that we've got to drive out.

CO-CHAIR ASPLIN: I would agree with Larry's comment, and you may be right, Joe. You may look at the --- you know, so I think there's reset opportunities that are today right in front of us relative to waste and variability that we need to capture. Now, how technology affects trend over time is an unknown factor looking forward. Joe, do you want to comment?

MEMBER STEPHANSKY: I just think when Andy is here tomorrow we'll cover probably more of the things that we've discussed in trying to get down to, for example, what Mary Ann was talking about in terms of hospitals don't know what it costs them to do this. Actually, I would disagree, we're getting very good at it, but we don't want to give away the store either.

The things that you're getting at I think we will get to tomorrow, but I would rather have Andy here to start talking about

them, and then he and I can bounce things off each other. He and I don't agree on everything either.

CO-CHAIR ASPLIN: Ariel.

MEMBER BAYEWITZ: Yes, I was just going to comment on the significant variation, just again on her point before. I think if that was stripped out, risk stripped out unit cost, I think if you add those pieces into the equation then you're going to get much more variations across the system. I'm just thinking within ---

CO-CHAIR ASPLIN: Great. One just real high-level comment I have. I like the diagram in that it shows some of the inner-related components of value and so forth, so from that standpoint it captures the contributors to value. What it's missing for me, and maybe this is captured in some of the work that the Linking Cost and Quality group has done, but is sort of a longitudinal timeline of where we are today, and where we're trying to

go. And, you know, I could see a couple of different figures capturing these elements over a time frame that might be helpful complements to this. One around total per capita spending, you know, what is the quality, experience, engagement, functional status journey from here to some point in the future? What is the total cost and resource-use journey, and how do they come together to create a better, more well defined picture of value at some point down the road? Maybe that's exactly what you did.

And then another longitudinal picture story of where we could go from a bundle payment perspective? Kind of envisioning the two at least obvious business models now that hopefully would get us closer to value than fee-for-service reimbursement. There may be others, but those two pictures might be a nice complement to this. Cheryl, do you have a comment?

MEMBER DAMBERG: I actually had an

earlier comment that was taken care of, but I like what you just said about this sort of bundling notion because I think that, you know, if we're trying to think about combining sort of everything that happens to the patient, that is in terms of resource-use and quality, you somehow or other have to get to this concept of a bundle. And I know that there's concern about all these episode groupers and the proprietary nature of them, but I think somehow or other we have to tread in that space if we want to get some measure of value going.

CO-CHAIR ASPLIN: So, we're going to wrap up. I don't know who still has an outstanding comment. I think Ariel, Taroon, I don't know if you had something. Do you have another comment, Ariel?

MEMBER BAYEWITZ: Just very quick just in terms of the image. I mean, we use it in one of our programs, you know, which is a Shared Savings Program similar to Medicare's

program where you have a total cost to care target, so a risk-adjusted PMPM target. And then you have sort of sliced, if you think about that as a pie, so your $400 average commercial for Medicare it's eight, nine thousand, whatever it is, and then within that there's a --- it's a pie and there are slices, and each of those slices are opportunities around resource-use. So, from a plan perspective, and we haven't talked about this at all, but thinking about leakage, thinking about lab being done in a hospital versus reference, thinking about surgery options, hospital versus ambulatory surgery centers, thinking about radiology, I mean, you could go down the list. There's all sorts of opportunities, and those are also resource-use. Right? I mean, we're not really talking about that, but those are examples and some of the stuff we're talking about, that they're all slices, we're giving directional slices. Ultimately, Dartmouth has their choice of following that advice or

looking somewhere else, but it is still helpful to cull out very specific opportunities within that global payment structure.

CO-CHAIR ASPLIN: I'll give Taroon the last word. Joe, do you have a quick comment before that? We're going to capture this again in the morning. I'll let Taroon get the last word, and then we're going to have --- see if there's any public comments, and then we'll do logistics for tonight and tomorrow.

MR. AMIR: Yes, I don't know if this is last-word-worthy, but I just wanted to provide some --- you know, we had a number of conversations over the day around measures of expenditure, or actual prices paid by the health plan for providers. And one reflection I had on the total cost of care measure that actually uses actual prices paid was that when that went through the endorsement process, and some of you will remember this. There was significant amount of push-back not only for that measure but also the ETG-based episode

measures that used actual prices paid down to the individual provider level, that there was a lot of --- there was significant amount of push-back from some stakeholders around whether, you know, individual providers or even larger provider groups can be held responsible a) for input factors, wages, differentials, things of that nature in their communities, and whether it was appropriate to be comparing regions based on these variables that may not be under an individual provider's control, if you will. So, it's an interesting question when we look at the --- even this question around the green box, which I agree with Cheryl in some ways that we're still working through this.

The way that we measure cost, whether it's episode verse total cost of care, I think that's an important differentiator. And then also the pricing model and the fact that standardized prices is --- the way you do the standardized prices is still --- there's still variation there. And if you use actual prices

paid using an expenditure approach, I'm curious to see --- you know, that was two years ago. Maybe the environment has changed and people are willing to, you know, engage in a conversation around the actual prices paid and take more accountability for that. But, you know, I think there's still some significant differences across our stakeholders around the acceptability of that as a measurement approach, particularly when comparing across regions. And it's just an observation to say that, you know, there obviously is still work to do in the green box.

CO-CHAIR ASPLIN: Yes, I can understand that across the country. It would be hard to justify it within a market. Sorry, I lied about giving you the last word.

Ashlie, can you --- let's see. We move to public comment, if there --- are there any --- Operator, if we could see if there are any public comments or questions.

OPERATOR: And at this time, if you

would like to ask a question or make a comment please press *1. There are no questions or comments at this time.

CO-CHAIR ASPLIN: Thank you. Ashlie, do you want to help us out with the logistics here tonight and tomorrow?

MS. WILBON: Sure, I'm actually going to let Lindsey do that. She's probably better than I.

MS. TIGHE: So, for tonight if you want to join us at an optional not paid for happy hour, it'll be ---

CO-CHAIR ASPLIN: I.e., if you want to buy Lindsey a drink.

MS. TIGHE: It'll be at Mio, which is on Vermont Avenue right across the street from your hotel. We'll be starting at 9 a.m. tomorrow with breakfast at 8:30 a.m., trying to wrap up around lunchtime, so if you need to check out, which I assume you will, there's space for your suitcases here. Just a small plug, if you have a little free time tonight and you want to read

the Linking Cost and Quality paper, it is available through the discussion guide. You can link out to it and download it. It will inform our conversation tomorrow.

I think that's it for me. Ashlie, is there anything else?

MS. WILBON: I think we'll email you guys the attachment paper, just to make it a little bit easier, so instead of like five clicks you've only got one, so we'll email it and attach it to the --- it's a great read for bed.

MS. TIGHE: For those who joined us on the web, it was a long day so thank you very much, we appreciate it.

MS. WILBON: Yes, certainly appreciate it. Thanks, Cheryl.

CO-CHAIR ASPLIN: Thank you all.

MS. WILBON: And thank you to our Chairs for getting us in early, and ahead of schedule. We appreciate it. Thank you.

CO-CHAIR LATTS: We're experienced

now.

CO-CHAIR ASPLIN: Okay, have a good night.

(Whereupon, the above-entitled matter went off the record at 4:54 p.m.)