NATIONAL QUALITY FORUM

Moderator: Diane Ferguson
June 6, 2017
2:00 p.m. ET

OPERATOR: This is Conference #: 94563908.

Operator: Welcome everyone. The webcast is about to begin. Please note today's call is being recorded. Please standby.

Ashlie Wilbon: Good afternoon everyone. Welcome to the Cost and Resource Use Standing Committee Post-Comment Call. Today, we'll be discussing the comments hat was received following the review of measures, three measures that we evaluated in March. We'll get into more detail what those are now. But thanks again for joining us.

This is Ashlie Wilbon. I'm a Senior Director here with NQF and I'm going to be helping to facilitate today's while – until Erin is able to join us a little bit later. I will take this opportunity to allow the other team members to introduce themselves and for the roll call.

Hiral Dudhwala: Sure. This is Hiral Dudhwala, the Project Manager on the Cost and Research Use Project.

Irvin Singh: Hi, good afternoon. My name is Irvin Singh. I am the Project Analyst for this project and I'm happy that you guys were able to come and join us today.

Suzanne Theberge: And this is Suzanne Theberge. I've joined the team as a senior project manager.

Irvin Singh: All right. So I'm just going to announce everybody's name, we're going to do roll call and just please say present when your name is called.

Do we have Brent Asplin on the line?

Ashlie Wilbon: He is calling right now.

Irvin Singh: OK. Cheryl Damberg?

Cheryl Damberg: Present.

Irvin Singh: Larry Becker?

Larry Becker: Present.

Irvin Singh: Mary Ann Clark? Jennifer Eames Huff?

Jennifer Eames Huff: Present.

Irvin Singh: Troy Fiesinger?

Troy Fiesinger: Present

Irvin Singh: Nancy Garrett? Andrea Gelzer? Martin Marciniak? Kristine Martin Anderson?

Kristine Martin Anderson: Present.

Irvin Singh: James Naessens?

James Naessens: Present.

Irvin Singh: Janis Orlawski? Betty Rambur? John Ratliff? Srinivas Sridhara? Bill Weintraub? Herbert Wong?

Irvin Singh: Dolores Yanagihara? All right, that's roll call.

Ashlie Wilbon: OK. Thank you, Irvin. So we'll go ahead and get started. I'm going to pinch hit until Brent is able to get in to the call. And just to give everyone a brief reminder of the measures we're going to be discussing today.

There are three measures that we reviewed during the meeting in March. And – 1598 and 1604 both of which are from HealthPartners or the total resource use and total cost measures. And 2158 which is the Medicare Spending per bene for the hospital level from Acumen and CMS.

All three – those measures were maintenance measures. We have reviewed those before and this was their maintenance review and all three were recommended for endorsement during the meeting in March.

So, next slide. So, in terms of the comments that we received – received a few comments, 21 comments are in the comment period from nine-member organizations. We did distribute a table of the comments received and staff has initiated responses to many of those comments on behalf of the committee which we'll be going over today.

Also for those comments – relevant comments we did also reached out to developers for them to submit their response and provide input for some of the comments that require the more in-depth response from them.

So with that, we have identified a few themes that we'll go over in a little bit more detail today. The goal today's call is really to review these major themes as well as the responses that were initially drafted by NQF staff to make sure that they reflect the committee's consensus and the response for those comments and identify any other issues or additions to the responses to the comments that would need to be addressed by the committee at this time.

So, many of these comments and themes are those – we have heard before given. These are maintenance measures that we've discussed many of these issues, many times (inaudible) in previous evaluations. So, none of these will probably seem very new or different to you but we did also – we did want to give the committee an opportunity to review them again and provide any additional input.

So the themes that we identified were around concerns for reliability and validity. The adjusting for social risk-factors, concerns about the populations include in the measures. There were several comments that were submitted and support for the measures. And also comments around updates to the Cost and Resource Use Measure Evaluation Criteria.

So I'm just going to do a quick check before we jump into discussion of the themes. Is Brent on the line yet?

Brent Asplin: Yes, I'm on. I've been here for …

Ashlie Wilbon: OK. Hey, Brent. OK, hey Brent. How you doing? Thanks for joining us. So, I will – I'm happy to do a quick overview …

Andrea Gelzer: I'm sorry, this is – this is – this is Andrea Gelzer, I'm on the line as well. Thank you.

Ashlie Wilbon: Oh OK. Hi, Andrea. Thanks for joining us. Has anyone else joined since we did roll call and started and kick the call off?

Nancy Garrett: Yes, this is Nancy Garrett.

Srinivas Sridhara: Yes, this is Srinivas Sridhara.

Ashlie Wilbon: Oh OK. I think I heard Nancy and someone else?

Srinivas Sridhara: This is Srinivas Sridhara.

(Crosstalk)

Ashlie Wilbon: Oh Srinivas – great. Thanks. Anyone else? OK. So, Brent, like I said, I'm happy to kicks things off unless you want to kind of get the committee going on the first theme or?

Brent Asplin: Why don't go ahead and walk through the (policy) group the themes and what your propose responses are and then I'd be happy to facilitate any dialogue among the committee.

Ashlie Wilbon: Sure, no – sure, no problem. So the first theme was around concerns for reliability and validity and more so measure specific comments both for 2158 and 1598. So for 2158 for the Medicare spending per bene, again, concerns around the association with measures of readmission in terms of identifying or establishing validity for that measure. And commenters also commenting again on their concerns around the post – spending in the post-acute time period driving the cost for that particular measure.

And for 1598 there were concerns around the testing for this measure. This measure is – have been developed by HealthPartners which is – in Minnesota and their testing was only in two states which included Minnesota and they're also requesting additional details around the standardized pricing, risk adjust – and risk adjustment approaches.

So, in terms of the responses to those themes, the staff initial – initiated were around the discussions at the committee has already had. And around these measures essentially filling that what the developer has submitted at this time was adequate to meet the NQF evaluation criteria and potentially around – especially around the post-acute period driving the cost for the possible base measures that this has been a topic that has been discussed by the committee as well. And that – there are – have been consensus among the committee that – while there are cost outside of the hospitalization that could be driving some of the results to the measure that there is an opportunity for hospitals to implement some practices to improve their cost for the measure.

So with that, I will pause and hand it over to Brent to see if there's any committee discussion on this theme.

Brent Asplin: Thank you. Do we have any comments on the proposed response or other comments related to reliability and validity on the measures?

Cheryl Damberg: Brent, this is Cheryl Damberg. I think that the committee response looks good and I think it accurately reflects discussion that the committee had. So I would support moving forward with that response.

Herbert Wong: This is Herb Wong and I agree with Cheryl's assessment in terms of the proposed committee response.

Brent Asplin: Very good. Thank you.

Andrea Gelzer: Andrea Gelzer, ditto.

Larry Becker: Same here for Larry.

Brent Asplin: All right, any concerns or suggested edits to the response for other issues under the reliability and validity of these measures?

All right, hearing none. We will have obviously an opportunity near the end before public comments to just revisit if needed but hearing none, why don't we continue through the deck and the other themes that were provided by the committee staff and the NQF staff.

Ashlie Wilbon: Great. So, the second theme was around Adjusting for Social Risk Factors. So again, within the SDS trial that NQF has currently going on, we do allow the developers to present a conceptual justification or basis for the risk factors that they've included and the risk adjustment model and the way the measures been testing.

And there were some concerns around the conceptual basis and justification for the risk factors, feeling that it was inadequate and that there were other factors that were not considered that are available in the literature that perhaps were not considered by the developers. And there was a request that there be a more in-depth look at the needs for SDS adjustment and the impact of these measures could have. And since that there's more testing that the community or the commenters would like to see around SDS factors.

So, this response is pretty dense. I won't read through it as you could do that on the slide or in this table. But essentially, the gist of the response is around, again, the committee's extensive discussion around SDS for these measures along with the discussion of developers at the meeting, and that there are still many limitations in the data. So while there are maybe conceptual links in the literature, often times finding the data to support those factors for testing is very difficult, and that the committee recognizes these challenges and for what was submitted that the committee felt that that was adequate and for the

testing that was done and that there will be additional opportunities to explore the variables and for additional testing. But that this time, this was what was submitted that the committee felt that the developer met that requirement this time.

So with that, I will pass it back to Brent in the committee for discussion.

Brent Asplin: Any comments on this topic from committee members? Maybe I'll start with the question. Well, these measures go through any of the same types of trial period and subsequent factors or discussions on sociodemographic factors or was that just with the prior phase?

Ashlie Wilbon: So, it terms of the trial, we are still in the process of determining whether or not the trial will continue or whether there'll be a permanent thing or whether we'll just continue with the piloting of that particular process, so that still being decided by NQF leadership.

However, if it is continued, if we do decide to continue with that when these measures would come back or meet us again, that that would potentially be something that would be evaluated again. I don't if that answers your question Brent of what you're asking.

Brent Asplin: OK. So TBD in terms of whether there's any subsequent conversations or we're going to come back together by phone to discuss anything, but it does, it sounds like it's somewhat different than the last time where we had a plan to work with the developers and review the conceptual model and review additional testing and so forth. At least at this point, that's not on the books.

Ashlie Wilbon: OK. I see what you're saying. Yes. So right now, no, I think it's the time that we did that. Last time we were just starting the trial and so we had already evaluated the measures without kind of the parameters of what we are looking for, what the conceptual link and the, yes, for the additional analysis. And so we added that piece on at that time.

So, this maintenance process kind of reiterated that and gave the developers an opportunity to come back with additional analysis from that initial review. But going forward, it would – any additional review of the SDS factors would

be a part of, you know, the maintenance of the measures or there was an ad hoc review or something like that. But there's nothing else planned right now beyond that.

Brent Asplin: Yes, that was my assumption. I just wanted to clarify that for the committee.

Ashlie Wilbon: Yes.

Brent Asplin: Are there comments about this theme around adjusting for social risk factors? Either comments from the developer or from the public comments or from the proposed response on behalf of the committee?

Jennifer Eames Huff: This is Jennifer Huff. I had a question. I think one of the comments referenced to that, the developer didn't use some elements that have been commonly used in literature. Couldn't find where that comment was, so I was wondering did the comment specifically identify which data elements they were talking about that's in the literature?

Ashlie Wilbon: I'm going to ask Hiral or Suzanne to try to …

Cheryl Damberg: Yes.

Ashlie Wilbon: … define that or respond to that.

Cheryl Damberg: Actually, to Jennifer, this is Cheryl. I had that the same question. And the commenter had submitted information. And the factors that they identified from literature were living alone, having unmet functional needs, lacking self-management skills, having limited education, inadequate health literacy, education levels, occupation, renting versus owning their homes and poor access to medical care.

Jennifer Huff: OK.

Cheryl Damberg: OK.

Jennifer Huff: That's helpful. So a good comprehensive list of elements, but it sounds like many of those so are challenging in terms of accessing the data.

Cheryl Damberg: Yes, I would say, virtually all of them do not exist in any data set that either Medicare or other payers would routinely maintain. So I think that that is sort of a larger challenge for anyone doing SDS adjustment work on any set of measures.

Andrea Gelzer: So this is Andrea Gelzer. We are starting to look at, you know, how can use E codes. I know that NAC has at a (pro-par) tools so they've got their practitioners trying to collect these various data elements. But I think until we all agree on at least a starter set of, you know, coding for some of the stuff, I think your comments are totally appropriate and we just – that's all we could do.

Brent Asplin: Yes, I think this topic clearly continues to come up and as we stated in the past, we certainly we open to new measures and new analysis as data become available. However, based on our conversations, I haven't heard any new information that would, at least drive me to a different conclusion relative to reliability and validity on this particular measure on the basis of sociodemographic adjustment.

Cheryl Damberg: And this is Cheryl. I would agree with that.

Herbert Wong: And this is Herb Wong and I agree with that assessment as well and I believe that the proposed committee response is adequate.

Larry Becker: Yes, this is Larry. Yes, I agree with that and there's certainly more work happening to look at – look at all the stuff and figure out the validity.

Brent Asplin: All right. Any other comments on sociodemographic adjustments?

Ashlie Wilbon: OK. I will go ahead and move on to theme three, which was around concerns about populations included in the measures. There were some concerns from commenters who work with cancer patients about how – about the variation and cancer treatment needs, the comorbidities and how about effect, cause and resource used.

And then another concern around the inclusion, particularly for 1598 and 1604 where OB/GYNs are included as primary care providers as part of the – one of

the specialties that is included, primary care providers for these measures, and the idea that some OB/GYNs are more specialized while some are generalist and the difference between – and the difference between those. And then concerns around providers being able to control those formularies and cost of medications that are prescribed.

So in terms of the response to these comments, I believe there's responses in the comment table for these. Also, I think these are issues that the committee has discussed particularly around the primary care providers and those that are included in the attribution model for or not even the attribution model, but included in – and who – which patients are counted in for primary care for the HealthPartners measures.

And I will – Cheryl, I don't know if it's – if it's easier to pull up the comments or maybe read the comment responses for these comments to the committee so they can discuss those, since it's now on the slide?

Hiral Dudhwala: Sure, we can – we can pull up the document. So we do the comment table then.

Ashlie Wilbon: Maybe if you could just give a summary or show the summary of the responses to those particular comments, so it will be easier for them to have a discussion around this.

Hiral Dudhwala: Sure. We'll pull it up right now.

Ashlie Wilbon: Those are responses that the developers provided responses to. So, I'm not sure that – that a detailed response from the committee is warranted, but it would be helpful just to kind of get an idea, make sure the committee has an understanding of what the developers responded with so that can determine whether or not any additional action is needed.

Hiral Dudhwala: Yes, we have it pulled up. Let us know if you all can see it.

Male: It's really small.

Sue Knudson: Ashlie, this is Sue Knudson from HealthPartners. I just wanted to point out that you were talking about the OB/GYN comments.

Ashlie Wilbon: Yes.

Sue Knudson: But what this theme three is the oncology response. So, you might want to go down to, I think, later in the document as the OB response.

Ashlie Wilbon: OK. Yes. I think those are both grouped into the same theme here.

Sue Knudson: Yes. And I'd be happy to make some general comments if it's helpful for you guys, but I want to respect your process.

Ashlie Wilbon: No, sure. That's great, Sue, that will be good while we're kind of getting the – something up on the screen for folks to be able to view visually.

Sue Knudson: OK. And then the team is here with me too as well. So they'll jump in if I've missed something. But generally speaking, so as a reminder, attribution in how we tested the measures base on what is accepted in this market and we've had market wide sort of adoption and input around our specification for attribution here for our youth, which is guidelines in the NQF submission.

We do include all services that are deemed to primary care. However, as it relates to OB/GYN in this particular comment, we're excluding and this is in our response. We do exclude certain subspecialties like GYN surgery from the attribution. So we're really trying to hone in on those that are more up to provide those primary care services, which has been done with the input from our market.

And then the other part of – it was the pharmacy comment related to this. The other part of this in our response around providers being able to control, you know, the cost of pharmaceuticals. We do see and believe that with pharmaceuticals at least in our book of business estimated at about 20 percent of total costs that it's really imperative that we have all hands on deck to help in supporting shared decision making around choices of pharmaceuticals.

So, putting this out there along with the improvement tools that we provide to help guide the best decisions for the patients and supporting their care teams with that information is important. And just like the rest of the measure, it's an all in measure that includes everything including pharmaceutical cost. And that's a real pressing issues as of recent because I'm sure you're all aware and so we felt pretty strongly it should remain included.

Those are sort of the high points. I guess the last thing I would add on that, which is not in our response, but we involve and engage providers and constituents consumers as well in our processes and we do have providers, for example, participate on our pharmacy and therapeutics that help drive those best choices. So, you know, we're very transparent in sharing our best practices and processes and pretty much have it all of that information on the public domain.

Ashlie Wilbon: Great. Thank you, Sue. So, again, this – well, that's – they were sets of comments around the pharmaceutical cost and the OB/GYN were pretty targeted towards the developers. So we did reach out to them for this – for response which Sue just nicely summarized for us.

So, basically, I just wanted to make sure that committee was aware and see if there's anything additional that the committee would like to add in response to that comment.

Brent Asplin: Hi. This is Brent Asplin. I would – I agree with Sue's comments and agree and support the including pharmacy given. I think given the rapid growth in pharmacy cost is just sort of increases the need to have as she put it all hands on deck, so I'm supportive about that. And the specifications around OB/GYN, I'm comfortable with their response as well. I think it changes any of my opinion about the measures overall.

Cheryl Damberg: Hi, (Jill). This is Cheryl Damberg. I definitely think pharmacy has to be part and parcel of this. You know, I've seen this measure operationalized in a number of different markets and if anything, I think inclusion of the broadest set of variables in the measure helps create dialogue among the stakeholder so that it has – they're going to us cost.

Ashlie Wilbon: Great. Thank you. So why don't we just kind of skip up to the cancer – or the comment related to the care of cancer patients driving cost. I know, Sue, you guys had to response to those as well. Did you want to do a quick summary of your response with that as well?

Sue Knudson: Sure. I'd be happy too. And, again, we'd just ask the team here to augment whatever I see at a high level. But first and foremost, we – in our response, I just wanted to articulate that we too are sensitive to the complexity of including cancer patients. However, the measure are – is a full population measures.

So to the extent that all groups caring for a population of patients likely have proportionately, you know, there's no like skew distribution towards one group or the other having that because the prevalence, also relative to all conditions in this commercial population would be relatively low.

So then, it sort of also relates to how the measure is being used, which is not really a part of this process, but what we'd advise is right now, you know, we're using it at the population base attributed to primary care. We're not attributing to cancer specialists, if you will, and realize and understand the need for more specific clinical information if you were going to use it in that manner.

What I'd also add is the measure does truncate cost at $125,000 as a reminder for those high cost patients. And then the N size, of course, at the population level would also help to ensure that a single patient who might have cancer or be an outlier would be skewing any of the results.

What I'd also say just from a usability point of view, we've had the measure in place in our market for years. As you all know, we've got over 200 users across the country. We've never had an issue come up in this topic area. So, I just wanted to add that commentary as well.

Ashlie Wilbon: Great. Thank you. Is there anyone from CMS or Acumen on the phone that would like to summarize their response to the issue around the care of cancer

patients on the impact on cost and then we can just open this issue up to the committee to see if they'd like to add anything.

Sri Nagavarapu: Hi. This is Sri Nagavarapu from Acumen. Yes, I'd just like to echo the comments from HealthPartners on the issue of cancer. We also – we're concern about cancer as well as other comorbidities that would tend to have very persistent effects that could affect episode spending.

At the same time, we didn't want to pursue sort of the wholesale exclusion of beneficiaries with these sorts of conditions because of the importance of making sure that the widest range of beneficiaries is possible or are within the measure and that hospitals have incentives for coordinated care of the widest possible set of beneficiaries.

And so, instead of going the root of excluding this population, we went the route of including the population in the measure and doing our best in terms of risk adjustment. And risk adjustment is quite detailed in the measure in terms of the (coverage) that are included in our response. We talk about the full list of (coverage).

The ones that are most relevant to cancer are several MS-DRGs that are related to cancer. So we include dummy variables to capture episode spending differences related to each MS-DRG as well as hierarchical condition categories, HCCs that capture cancer.

And so the risk adjustment models have both HCCs and these MS-DRGs and they're estimated separately by MDC with the hope of providing a flexible specification that can capture the influence of the sorts of comorbidities that we don't want to assign the consequences of – to the hospital in the measure. And I'll stop there.

Ashlie Wilbon: Thank you. Those are very helpful. That's maybe just open it up to the committee and see if there's any other additions to what the developers have shared and response to that comment.

Brent Asplin: Yes. And so it's helpful to have both developers on the phone, appreciate that. And just to ask for any additional comments, questions or input from the committee relative to either of these populations.

James Naessens: This is Jim Naessens. If I recall right, the HealthPartner's measures do not actually have a specific attribution methodology, but they're submitting guidance for that. We may want to reflect in our committee conclusions that we recognize that the specific attribution was outside of our venue to decide.

Brent Asplin: Thank you, Jim. I guess the only question I have – I mean, there's always limitations in risk adjustment and obviously the risk model in the captures of several to MS-DRGs that are relevant to cancer on the MSPB Measure 2158.

And, you know, I guess the concern from the committee could be if you have disproportionate percentage of those patients is that adequately captured by the risk adjustment model. And I don't know the developer has any additional comments on that. We know it's not perfect, but certainly the intent of the risk adjustment models to capture those differences and if you have any comments about when the ability of the model captured may become a little weaker.

Sri Nagavarapu: Yes, thanks for the question. This is Sri from Acumen. We then testing in general on predictive ratios overall by deciles to kind of make sure that expensive cases regardless of the source, so whether it's cancer or from other conditions are being adequately captured by the risk adjustment.

So, I think in terms of empirical analysis so far, we don't have something that specific to cancer, but we have something that's broader that I think it's at the underlying concern about whether particularly expensive cases are being adequately captured in risk adjustment and those are the predictive ratio results.

The other thing I should mention in your question reminded me is as in the HealthPartner's measure, we have an inclusion related to outliers. So, we calculate cases where the risk adjustment model is not adequate to produce reasonable size residuals and look for particularly large deviations and exclude those from the measure to deal with the smaller set of cases where it

seems like the risk adjustment is not able to capture very, very extreme spending.

Brent Asplin: Thank you. Any other questions from the committee members relative to high class cases and summaries, especially, specific to population?

Kim Spalding Bush: Hi. This is Kim Spalding Bush from CMS. And my question is actually two or three. I don't know if it would bear mentioning in this context and I don't know how detailed the response even gets. I've not been through this process recently. So, is it worth mentioning that if we do get to a place where the episodes get to the very highest?

We do address those kind of high cost outliers as well. So I think – like as you said, the risk adjustment, you know, generally does do a good job at predicting cost than any event that was not found that are outliers. The measure construction outside the risk adjustment does account for that.

Sri Nagavarapu: Yes, that's right, Kim. We do exclude outliers base on the residuals from risk adjustment. So, in cases where there's extremely high spending episodes that aren't being captured with the risk adjustment model, those cases will be excluded from the measure. This is something we could definitely include in the comment response, Kim.

Brent Asplin: All right, thank you. If there are any other comments or questions or suggestion we keep going with the deck.

Ashlie Wilbon: Sure. So, theme four, I won't spend a lot of time on it. It's just around – just recognizing that there were several comments submitted and response to all three measures and support for continuing to endorse cost measures that this is a gap in measurement for this area.

So, our response is really just acknowledging the comments and thanking the commenters for their comment. I'm not going to pause there unless anyone has any objections. But let's go ahead and get to theme five, which was around Updates to the Cost and Resources Measure Evaluation Criteria, which we discussed toward the end of the meeting in March. And there were some commenters that were asking for clarification around the performance gap

subcriterion. And this is the response that was primary initiated and response from NQF directly.

And basically acknowledging the comments and that the purpose of the performance gap subcriterion is really to determine whether or not there is a cost problem around that particular topic area. And, you know, acknowledging that, you know, NQF wants to endorse measures that where we know there are problems in measuring cost or whether are higher cost than there's a need to better understand what's driving those cost within that particular topic area.

And so, really the changes that we made was to the subcriterion with the streamline the criteria and ultimately help to prevent any redundancy or talking about the same issues when we get down to the reliability and validity subcriterion.

So, hopefully that response was adequate and clarifies that for the commenter. But I don't know if any of the committee members had any response or addition to that but again that was one that was formulated directly from NQF for clarification.

Andrea Gelzer: Excuse me, this is Andrea. Are there other measures? One of the comments suggest that, you know, that this measure is based or the HealthPartners measures based on the using ACG's as a grouper. Are there other measures that you are aware if they are under development or in use that will be coming to this committee that have other grouper?

Ashlie Wilbon: So, we're actually going to be talking about that as a part of – kind of the next portion of this call. In terms of the risk adjustment approach using some sort of grouping tool, I'm not aware of any and I can open that up to – I know – I think Taroon is on the phone he may be aware. But I don't know of any other that we've endorsed so far that used of similar risk adjustment tool like ACG.

Taroon Amin: Ashlie, this – Ashlie, this is Taroon. Yes, you're correct.

Ashlie Wilbon: Yes.

Taroon Amin: I mean we don't have any endorsed measures.

Ashlie Wilbon: OK.

Taroon Amin: But other folks are using, you know, the HealthPartners measures and other measures using other risk adjustment method or other risk adjustment grouper methodologies. But, yes, to your questions specifically, no, there are not other measures.

Andrea Gelzer: So, I'm fine with your response.

Ashlie Wilbon: OK. There is another section, we will talk a little bit about some of the measures specific comments that came through where I think that risk adjustment approach particularly for the HealthPartners measures theme came through there. So there will be potentially be addressing that again. But we will be having a separate discussion outside of the comments after review the comments about the episode grouper-based measures that maybe coming for in the future. So, more to come on that.

Any other comments around the updates that we made to the resource use evaluation criterion in spring?

Male: Somebody's on here.

Female: Oh, I'm sorry. Grab the phone. Are you on the phone?

Male: I was (Linda).

Female: I'm sorry.

Ashlie Wilbon: Hello. Hello?

Cheryl Damberg: This is Cheryl. I don't have any further comments. I think the committee's response is clear.

Ashlie Wilbon: OK. Great. So with that, I believe in next section is for us to do an overview of some of the measure specific comments that came through. And most of these comments were directed to the developers to provide responses, the

opinion or memo that they both provided pretty extensive responses to the issues that were brought up in the comments around the measures.

So, I would actually like to give them the opportunity just to summarize for the committee again what their responses were for this particular issue. Some of them were addressed in previous themes. But if there's anything that was not addressed, I'll start with HealthPartners. Sue, if you want to just provide the committee of brief overview of your response to some of the measure specific themes that we didn't already discuss.

(Off-Mic)

Ashlie Wilbon: Go ahead, Sue.

Sue Knudson: OK. Yes. I'm actually, you know, we tracked, we read through how you've seen the comments. Thank you for your comments about us writing through responses, we tried to be. I felt there were themes hit the high points of the categories of them. So, I'm not sure I have additional comments. Within comment on the first two themes, you guys seems to go through those OK.

What I would say on the lack of adjustment for social risk factors. You know, we landed squarely where you had summarized if we would have had those data available to us that was mentioned in a literature as conceptual topic ideas, we certainly would have tested it.

I also think in a full day meeting, we went through pretty thoroughly the results of our income-based study and also that the measures are limited to a commercial population in some of those items. So, I guess those would be additional comments around theme two.

And I think there was just one other thing around the measure testing and usability about being tested just in a two states. But we had also mentioned in our response that we've got several users. You know, over 215 across the country as well as in within the State of Minnesota, Minnesota community measurement doing state-wide reporting and other collaborative in other states. So certainly, the usability is there.

And then I guess my final comment will be on the support for the measure. Totally agree with your brief comment and thanking those folks for those comments. But I think it goes to the industry and those of us working in it being ready for the measurement to the full Triple Aim including affordability, and the fact that there is more letters of support than letters of concern. I think is very telling.

So, I guess that will conclude my comments.

Ashlie Wilbon: Great. Thank you. And I believed there was also some measure specific comments or the Medicare spending per bene measure, which I think may have been addressed but I do know there were some comments around testing the measure of the clinician level before adoption at the – in the MIPS program, and just encouraging the developer to move forward with testing for that. And I know that more or so at a problematic level but wanted to give the developer an opportunity to give us summary of their response for that as well.

Kim Spalding Bush: Hey, this is Kim from Centers for Medicare and Medicaid Services. I think we did provide a written response, kind of a high level that we developed in conjunction with the MIPS team, basically saying, you know, they are still working on the measures and sort of the future for the MIPS program. Isn't something that we feel is germane to this discussion than it sort of something that we can't comment on at this point either.

So I think those comments aren't something will be able to respond to you as a part of the endorsement process, if that makes sense? And since this is really specific to the hospital measures, so I think validating it at the clinician level for MIPS is out of the scope. And we don't even have the folks on the phone for the MIPS team to address that.

Ashlie Wilbon: Sure. Thank you.

Kristine Martin Anderson: Hi. Sorry, this is Kristine. I just – I think I read that part of the write up too. And I only have like two like minor suggestions. One is that there's a reference to the MIPS measure and I wasn't sure whether or not there was some ability to help people differentiate between them using some kind of

measure ID or something because they just referred to by this setting and says that it's still in maintenance or whatever.

And the two-time CMS answered the question. The paragraph was slightly different and the second paragraph was more complete. So just I don't know if the first one is even possible if there's a measure ID but for the MIPS 10 measure or MSPB 10 measure.

Kim Spalding Bush: So, we do have the – we could use the NQF number. I think that's a good suggestion. Where we attended distinguish the hospital measure. I don't know if CMS issued a Measure ID for the MIPS version of the measures. So I'm not sure if can do a number but we could put the NQF where we're referring to the hospital version.

Kristine Martin Anderson: Yes. And you are – that would working in your – it was in response three that you talked about – that the MSPB 10 Measure was finalized for inclusion and its cost categories as part, you now, the QPB final rule. That's just not in the previous answer. But I think it is helpful because it drives home the – their two different measures part of the discussion.

Kim Spalding Bush: OK, thanks. Yes. So, I think Acumen team can update so that the last complete response that contained that information as well. And then I guess – I guess they know the next step, I mean we get to the – at the end of the call for how to submit the revised responses back. Thank you for that.

Kristine Martin Anderson: Sure.

Ashlie Wilbon: Are there any other committee comments or questions for the developers around the developer – I'm sorry, the measure specific comments and the responses of the developer submitted?

So, again, I think many of these issues were discussed or kind of rolled up into the themes that we've already discussed around SDS, high cost of the cancer patient, the MSP – MSPB 10 measure we just discussed. I already said SDS.

So, I think a lot of those were already discussed in prior themes. But if there's anything else that any of the committee members have questions about in your

review of the comments or responses, now would be a good time to express those.

OK. Hearing none. I believed that concludes the discussion of the comments. Let's go ahead and open it up for public comments. Is there anyone on the phone who would like to raise an issue or questions to the committee?

Operator: At this time, if you would like a public comment, please press star then the number one on your telephone keypad. Again, that's star one to make a public comment.

No public comments at this time.

Ashlie Wilbon: OK. Thank you. So before we move on and transition to the next section of the call today, I just wanted to thank the developers again for providing – for joining the call and providing input on their responses and giving us summaries and being (above) in the call today. Any final thoughts on the comment and responses from committee members before we move on to the next section of the call?

Male: All good.

Ashlie Wilbon: OK. Let's go ahead and push forward and we may be able to get off this call, be able to finish up a little bit early, we're ahead schedule so that's great.

So the next section of the call was really to follow up on where we – actually we never really got to at the meeting in March and staff really wanted to provide you kind of circle back and provide the committee with an update of where we were with potentially some additional work that maybe coming down in the future cycles of evaluation for episode grouper based measures.

And then just to bring back some of the work that we had done before around linking cost and quality to see whether or not there's any action around implementation specifically for the operational guidance that was offered at the time we did that work.

This issue of the relationship of quality measures to the cost measures that we review continue to come up and we though it was – it would be a good time that kind of revisit that issue and see whether or not there's any things we like implement into – or integrate into the evaluation of cost measures that might further kind of push that or advance that – the cost of putting the – encouraging linking cost and quality measures together. So, we'll spend a little bit of time on that as well and get some of the committee's input on where we might move forward from there.

So we'll start up with the episode grouper work. So, I think some of you actually on the call were part of this work. I think Kristine was co-chair for that committee, I believe. So feel free to jump in at any time but just a quick refresher on episode grouping.

So this slide is one – is an image that we developers part of that work to try to illustrate what an episode grouper does. It is very technical software that basically takes claims and parses them into similar episodes. It uses grouping logic to – or decision logic to decide where the claims go into which episodes. And, again, this is largely done in software or primarily done in software and has been in used predominantly by commercial users at this point.

However, as we all know, CMS has been working on a grouper as well, so it is moving into the – to the public sector and we will – we are expecting potentially from episode-based measures from the public grouper at some point.

So just kind of looping back to the image. So, once the grouping logic has been applied within the software, you end up with several episodes which have group claims together that have been related in some way to a procedure to a timeframe, a particular treatment of a particular condition.

And then those episodes are generally used to develop measure where risk adjustment is applied, costing method is maybe applied. There are some sort of scoring mechanism that is developed in that kind of – is where the measure itself comes together.

So, essentially, what we are expecting to come forward at some point are these episode-based measures for evaluation by the committee. So we had episode grouper project back in 2014, where we were really trying to get an understanding of what it would mean, first of all, let's define what a grouper is, kind of expecting that this work will coming forward. We wanted to have some foundational understanding and consensus around what it would be that we would be evaluating and what that would entail.

And so part of that work the committee works on establishing definitions and foundational principles for defining, you know, what a grouper is, how it would evaluate or should be evaluated.

And one of the core principles that I pulled out that will kind of loop back to is around the need for potentially multiple phase evaluation, given that the grouper itself has a set of logic and rule that is necessary to understand how the episodes were then developed and then how the measure was been developed from the episodes. So they're all very closely and tightly link.

And the consensus from the group was that it would be really hard to just evaluate the measure that came from the grouper without understanding how the episode was establish and then without understanding the logic and the rules that were develop to – within the grouper to have the output of the episodes and there for the measures. So with that, really recommending that there would kind of multipart evaluation when these measures will come forward so that there is a comprehensive understanding of the entire tool.

They also laid down some foundational recommendations on what information will be needed in order to be submitted or an expert body to evaluate episode groupers. How and which evaluation criteria should be applied at both groupers and the episode level, and provided some initial guidance on how we should go about evaluating the groupers when they come in.

So again, the recommendations were around this multiple-layer evaluation process where NQF staff would be involve, there would be technical experts, potential clinical experts or evaluation of the clinical episode. And then,

obviously, some sort of multistakeholder committee input of the episode and the measures as well.

Given that, the expectation was that the evaluation of groupers would be pretty extensive, it would be something that we had never really done to this level before that we had taken this to our CSAC committee and recommendation was that our initials (were to array) into evaluating. Episode groupers would be more of a peer review process and not speak necessary to endorse them but to kind of go through the motions of evaluating providing input and getting some practice, if you will, for the committees and understanding how these things work before we step out there and provide the samples of approval on there and allow them to be specifically the public grouper and use a little bit longer before there was an endorsement designation put on there.

And then there was a desire for – strictly from the CSAC for us to use this effort as a lesson learned or pilot effort if you will to understand what some of the challenges would be and this initial effort so that we could roll that forward and potentially establish more permanent profit to evaluate the burden, group of base measures in the future.

So, I'm going to just pause there and see if there are any questions about that because I will quite mouthful just to make sure were all on the same page about groupers and some of the recommendation.

Andrea Gelzer: So, this is Andrea. Is that consistent with the CMS timeline for use of groupers?

Ashlie Wilbon: So, I'm not sure I could totally answer that question but I will maybe – if there are other staff members who have better insight into what maybe communications of CMS on what – where they are with the development and we'll get to that on a couple of slides. But my understanding is that there are few measures ready from the grouper. Few episodes, maybe seven or eight episodes that maybe ready around various conditions.

There was an expectation I think if you guys remember that some of those would have been submitted for the phase of work that we just completed but

the timing wasn't quite right and they weren't able to get those and for this process. So, the expectation is that, the next effort or next evaluation cycle that we do that we would try to time it so that CMS would be able to submit some of those measures for review.

Kristine Martin Anderson: Hi, this is Kristine. I'm going to comment on this CMS timeline directly but I just so compel to say something about our own expectation around this because, you know, imagine its 1980 something and your hand at the DRG system and told evaluate this and say (it's been) good.

The truth about a lot of these types of programmatic things where – to parse claims et cetera is that, what tells you how good they are is use. So I think being able, you know, it's (not) possible for like the committee to parse through all of the rules in a grouper and say, yes, I agree with this one.

And I think it's going to be create great episodes in these quality measures on top of these episodes will be great. So, I think it's important to kind of get a point of view around what are some of the characteristics of a grouper and its, you know, perceived advantages. And how does it looks like if applied and some small, you know, usually the measures are start out in a topical area.

And then, you know, have a healthy attitude about what, you know, use will teach you about how you would refine it with coding changes. And as you know once you start sorting claims in way that the public knows about these are changing the way they code to optimize their grouping and so it's like – it's iterative process to get to an end state.

Ashlie Wilbon: Right. Thank you very much for that. I think I second that. It is very much going to be an iterative process not only with the used process but with the evaluation process as well. So, I think you – we only get into few slides but there still a lot of I think internal thinking and work that we need to do here at NQF to really prepare for this and kind of, you know, educating the volunteers and the committees and experts that will be helping us view this effort.

So, definitely a lot more work to come and a lot thinking about this even though we did actually establishes a committee that Kristine led to help us, you know, initially get some thinking around this topic. There's still lot more

to do. And so any input you guys have to do would be great, again because we are still, you know, continue to think about this and prepare for the work to come.

So, let me go ahead and move on because I think I've covered some of the stuff. And again this is very high level over view. The report that we did for this project was a lot more detailed than highlighted some of the issues that Kristine just mentioned. So, I will just kind of keep plug in and please feel free to jump in and ask questions as we go. And this won't be the last we'll discuss this. There will be plenty of opportunity for us to go and do deeper dive on the report and kind of updates on where we are as the thinking evolves and as we prepare for the measure coming forward.

Next slide.

So again, some of the anticipated submissions have been measures that have been reviewed or discussed in the measure application partnership or the MAP. Again, there are – Erin hasn't quite joined us and she's got a much better list than I do. The high level is that there are about seven conditions specific measures that we know that have come through hospital-based measures that we know that have come through MAP that we expect could be potentially ready to come forward.

One of which I believe maybe a hip replacement measure, another one on potentially cellulitis. There's also potentially a – there are potentially measures that maybe coming through – that we might – may want to use in MIPS. I think there's 40 in the rule. I'm not sure when those maybe coming or what the topic areas are on those. And I think there's still a lot of communication and discussion we have to do with CMS on their readiness for some of these measures to come through. And how many of those will actually be from the grouper but I think we'll pretty sure at least seven to eight of them will be grouper based base on that summary.

So some of the anticipated challenges, as I mention, we're still thinking through many of this but obviously the political implication of the intended use of the measures, you know, for MIPS or other hospital-based programs,

balancing the developer burden with the submission requirements. So, again, I mentioned that kind of a three-phase evaluation. So, documentation so that there can be an understanding of how the grouper works, how those episodes were develop once the grouper logic was replied, and then the actual measures that were used with the episode as a foundation for those.

So, that potentially could be a lot of documentation. We did get guidance in the report from the grouper committee or expert panel around the submission requirement. So, honing that and just making sure we have a healthy dialogue with developers about what is actually necessary to submit so that there can be a good understanding of what's going on with the grouper in the measures.

And then following on to that, the volume of measures and the material that would need to be reviewed which then, you know, the (strap on) staff and potential (strap on) staff and volunteers to review the information, and the timeline that would – that the measures would be needed and run through our process.

Again, the initial recommendation from our CSAC was that our initial effort around reviewing episode-based managers and episode groupers would be not to endorse the measures initially. However there is some desire as we've heard kind of talking to users and so forth, 48 measures to be endorsed so I think there are some ongoing discussion that needs to happen internally and with our users around the importance of endorsement for this initial effort for reviewing the grouper base measures.

Again, coordinating with the multiple expert bodies potentially if we have technical experts, multistakeholder committee which would likely be the Cost and research Use Committee or other clinical experts depending on these clinical topic areas that the episodes are based on.

We do have several – we have 20 other committees – standing committees that have been established. Many of them are around specific clinical topic areas. So, the idea would be to lean on those and use those people as much as possible. They've already been pulled into committees but kind of coordinating all of those efforts potentially could be a challenge as well.

And then just to challenge us around implementing kind of a new evaluation process, around a new measurement construct and getting everyone up to speed and the education that would be required for the multiple layers of evaluation that will be needed.

And then, as many of you know, we've just completed a kaizen which was aimed at improving our consensus development process and trying to streamline our process or endorsement. So that process were currently internally and working – internally working and with some external stakeholders on streamlining the process and developing some new resources and hopefully improving our infrastructure to some degree, and so all of that will be going on as we're figuring this out as well.

So, lots of moving pieces and we are not at all misguided about how – about how challenging it will be to do this evaluation effort.

I'm just going to pause there and see if, Taroon, if you had anything to add to that or extend upon.

Taroon Amin: I don't have anything else to add. I mean, certainly, the challenges are many but certainly it's important endeavor as we kind of know these groupers are going to be use for various important applications both on public sector and obviously they've been used in the private sector as well. But, this would be the first step as Kristine pointed out and this is the approach that we're thinking through in terms of next steps.

Ashlie Wilbon: Thanks. So next slide. So, I've mentioned many of these points already but our ideas and our kind of approach to address any challenges really giving as much upstream collaboration and communication with the developers including CMS and any other who may have grouper base managers that they want to bring forward and just making sure there's an open line of communication and setting expectations about, you know, what the process will be and of what to expect for timeline and number of measures and so forth.

I think the other thing that we're looking to do is making sure that we have a realistic threshold for the number of measures that can be reviewed in a

particular process. I think, you know, or evaluations like, well, I should say I think, you know, there maybe some expectation that, you know, I only mention eight measures, maybe there's more and there are some desire to have more measures review, and just having a realistic expectation about what can actually be done by volunteers and staff being the first effort with all the other things going on for this particular effort and setting that threshold and communicating that to CMS and other developers.

I've mentioned training and education for volunteers, staff that includes developers, building the infrastructure so that the submission process, evaluation process can be a service as possible. And then again, you know, making sure that we're tracking how the process is going and evaluating it as we go so that we're ready make any improvement as we move forward for future efforts to evaluate the groupers.

Next slide.

(Mike): Hey, this is (Mike) from CMS. I have a quick question.

Ashlie Wilbon: Sure.

(Mike): Do you need our developers for the rest of this portion of the call or can we release them?

Ashlie Wilbon: Oh yes – no. Sure, you guys are free to go. We shouldn't need any other developer input unless you just like to listen what else is going on.

(Mike): OK, thank you.

Andrea Gelzer: And excuse me, this is Andrea Gelzer. And, I mean, and forgive for making this comment perhaps and I don't want us to go off completely (in) tangent. But that said, if we look at the resources that we have to do this work, I mean groupers to me are not ready for primetime. And I know that the land has deemed them an alternative payment model. So, it's a way to count alternative payment and to, you know, qualify for additional opportunity.

But if you, you know, if you think, OK, getting even, say you looked at cellulitis or you look at hips and knees or you look at what, you know, cardiac. To me, it's still a fragmented – again not ready for primetime effort. And I think, yet, I think we should further investigate it academically. But unless somebody can prove that by moving into bundled payments and grouper methodology to reward providers. Unless that's an evolutionary step to getting to a total population-based payment methodology that make sense, I guess I agree with the CSAC for other, you know, perhaps not quite the same reasons that were expressed but that we've got too much else on our plates to be worrying about.

Ashlie Wilbon: Any other comments from the committee or responses to Andrea's concerns?

Kristine Martin Anderson: This is Kristine, something I would say as it there are already bundle payment models in used and starting to get some experience with evaluating them, it seems like if you've ignore it entirely, then you miss a form of stakeholder and, you know, trying to get consensus input even around some of the steps where we are today.

Cheryl Damberg: Yes, this is Cheryl Damberg. You know, I concur with Kristine's original comments which is, you know, we have to view this since kind of a work in progress. And I guess my feeling is that the sooner this committee starts to digging into the space and kind of learning by doing, probably the better because I do think that this is where, you know, the payment models are going.

Brent Asplin: Yes, I agree with Cheryl. This is Brent. A couple of general comments, one, I agree with Andrea that the population-based payments, like the HealthPartners payments which were specified for commercial.

We do had – we did have a total per capita spending measure that did not reach consensus in the committee if you recall a couple of years ago or three years ago. And I do think we need a total per capita cost measure in the Medicare population and recognize that that's a different area of CMS and they're not able to respond to that today. But I do think that's a gap. And I

agree with Andrea in general that total per capita cost spending measures are something that we need to have in place for us the board.

As it relates to groupers and episodes, though, I would second Cheryl's comment that – I am a little concern for making this too complicated because they're already in existence. And as you see, the comments we've received and talked briefly about earlier today, there is fair amount of angst in the physician community about the resource component of MIPS. And I think there are already episode or a bundle of payment models for some selected specialty that will be far better than a clinician-specified version of the Medicare spending for beneficiary measure that we talked about earlier.

So, I think the 2158 MSPB measures specify at the clinician level is probably not the right answer for the resource component of MIPS. And so there's a time urgency here and to the extent that we believe in NQF endorsement for measures that are news and I don't think we'd be volunteering if we didn't. Then I do think there's some time urgency.

And then specifically, I think we've created a gauntlet process here for evaluating episode groupers and I would urge NQF that they really feel that grouper tools need to be evaluated separately that you quickly convene some sort of tap to go through that and create a list of usable tools.

The analogy of it use is that as outlined here, unless I'm misunderstanding it which is always possible, we would've had to evaluate. For example, in the HealthPartners measures, we would had to a comment and evaluate on the Hopkins ACG tool in addition to (evaluation) of the measure.

And I just don't think that's practical given the time constraints for these committees. And, frankly, some of the expertise that you would need to technically evaluate the grouper tools is going to be different than the expertise you would put on a consensus committee. So I am hesitant to separate. If you need to do that, I would just do it separately and have an approved list.

And, you know, at the end of the day, some of these episodes aren't very difficult to define, particularly, around procedures. I get that they get more complicated when we start talking about episodes for chronic conditions and that's where it gets much more grave. But one of the early ones, around acute care episodes are pretty straightforward and I'm worried we're making it too complicated. And by the time they're endorsed, it will already be several years in the MACRA

Andrea Gelzer: So, this is Andrea. So I'm hearing the – it makes sense to go forward. If for nothing else then to see that these measures do get tested to see if they are truly reasonable to continue with in the long term. I wasn't very articulate but is that what – does that make sense?

Brent Asplin: Yes, this is Brent. I think that makes – personally, I think that makes all the sense in the world. Yes. So it bothers me that the total per capita in Medicare spending that we didn't endorse because in use in the value-based modifier and will probably roll in to MACRA as a resource measure even though it doesn't have NQF endorsement.

Cheryl Damberg: Right. Yes, I mean I think I agree with this space, they look to us, I think we have to forge ahead.

Female: So does that change for – with regard to NQF staff and plans, any – the plans and resources that you're going to put on groupers?

Ashlie Wilbon: This is Ashlie. I think Brent's comments were very well put and very well taken. I think that, you know, we still have – we have – we did actually have recommendations for a TEP to review to the actual grouper portion. I think how we operationalize that, I think you're talking about how to do that and maybe (having that be a) separate effort and then maybe their work trickled down to other groups or group that reviewed the episodes and the measures or something.

I think that's something we need to continue to think through particularly with the timing of – when the measures will be ready, when the groupers ready and when we can actually start work. So, I think there's still a lot more thinking to do there. And we're really trying to figure out how to as much as possible

spread the evaluation across various types of experts so that no one group is tasked with doing everything. And maybe different types of expertise that is needed to review different types – different parts of the grouper, the measure, the episode or what have you. So, I think all of that is well taken. I think we just have to do some more thinking about how that would actually be put into action.

Kristine Martin Anderson: You know – it's Kristine again. Something I want to remind our NQF staff at least when an episode grouper committee had the discussions. There still were big philosophical divides even in the expert panel, right, that we didn't try to reconcile. For example, the biggest one being do you build an episode around the patient's experience or do you build and episode around the physician's experience.

And there are valid reasons to argue for both depending on use. And, you know, and what you're holding accountable for and what the payment model is, et cetera. I think that's the main reason why I said – we said, "Hey, you got to look at the whole thing." That doesn't you're evaluating the scientific validity of a grouper's software system, you know, every time you're looking at measure, but understanding the underpinnings of the grouper as it's being applied to a measure and then knowing what the measure is going to be use for are sort of critical.

So, I think it's there – you can't get to a best in class on grouping. That wouldn't – that would be a full (barring) right now. But you can at least build a set of people in the NQF community with knowledge of how these groupers work and what the tradeoffs are. Just like we've been doing around to socioeconomic status and, you know, started to do a special determinants of care. So, just trying to think about – it's going to be part of our future.

Ashlie Wilbon: Thanks. You know, that is very true. We did – I do remember that debate very well around the patient-centered versus provider-centered grouper, so yes, definitely more thinking around that. Taroon, any thoughts to add on the groupers before we transition to (looking constant) quality?

Nope, I think he's no longer with this. OK. So, that is basically – I want to thank everyone for their thoughts. And like I said there's still a lot more to come on that. Again, this is just – the purpose of this was really just to kind of bring it back in the forefront of your minds, let you know we're thinking about it. Some of the issues that we're grappling with right now and we may very well be coming back to you in the near future with just getting your input on a more concrete plan about how this will roll out.

Just a couple of things on those previous slides that I just wanted to bring up, some of the main issues that we're going to be looking at is kind of balancing this desire for endorsement in this initial effort when we may not feel like we're quite there yet. And, you know, just getting a sense for how different it will be to evaluate a grouper-based measure versus other resource-based measures. I think that's something that may be useful to have the committee discussing more detail in the future.

And then Kristine mentioned just alluded to this earlier but this issue around burdening and maintenance in how tweaks to groupers, in particular, are made very frequently based on user experience and with any other type of measure where, you know, tracking any updates and we revision the measure, the measure gets maintained. And when there's a measure that's based on a grouper that's constantly potentially – frequently and constantly changing how we track that and still ensure that the – and if eventually we endorse the grouper-based measure, how that measure is as consistent as possible for all people that are saying they are using that endorsed measure. So, lots more thinking to come on that and we will keep you guys informed on our – and how that evolves.

So, if there aren't any other thoughts on that, I'm going to jump into the linking cost and quality piece. OK.

So, this was, again, another project that we did back in 2014 to explore this – the issue around how cost and quality measures should be linked to measure efficiency. And the project was really centered around trying to identify some of the key (methodological) challenges the linking cost and quality, trying to identify some of the best practices and principles for doing that. And also a

desire to have some operational guidance on, particularly on NQF, on how we might implement some changes to our process to further advance the field and doing this link on (February) in the endorsement or at the programmatic level.

So, the project had kind of three major components. We had an expert panels and we also have commissioned officers that were kind of guided by the expert panel and they wrote a paper. In the paper, they performed the literature review and also described the environmental scan they did where they basically (done) an inventory across the country as all the different programs, and entities that are using cost and quality measures together, and the approaches that they are using to make this link either to display the measure or to develop some sort of composite, and to, you know, really make that link between the cost and quality measures more intentional.

So, through that scan, they identified several models that are being used, currently in use, or that have been mentioned in the literature. And then they used CMS dataset from, I believe, the hospital value-based program to compare performance results when you put the same data into different models and how the performance differ – the performance results differ based on how you link or what the relationship is between the cost and quality measure for that particular model.

So, next slide.

So, some of the key findings that they had is there were seven different models that they identified. I'm not going to go into a lot of detail about those but I can share those again with the folks after the call if needed. But there are potentially seven models that they identified in the literature and through their environmental scan. And basically, after doing the analysis, identifying the models and doing it now with the CMS data, you know, the key findings were that used case matters. So it matters what – how you're going to be using the cost and quality measures together, the application that they'll be used for.

It does make a difference in how the performance results come out and there were some tradeoffs. So, what are you group provided by quality first and then compare on cost within particular groups of – or threshold within the

quality performance or if you rank based on cost and then compare it on quality. So, there are different tradeoffs in the approach and based on how you're going to be using the analysis. But has to be clearly – there has to be clearly defined cost and quality signals as well as defining their relationships to each other in the model that you're using in order for you to make an accurate conclusion.

So in particular, one of the things that were identified is making sure that a time period or let's say if you have cost measure and five different quality measures around that particular topic area that the domains within those measures are aligned (in the) time period. All the measures are measuring cost and quality around the same – across the same time period and then they're capturing – all capturing the same measure population.

You can imagine with those things don't align the signal that you're getting when you have all those measures in a particular model may not necessarily be clear on exactly what type of efficiency or where the efficiencies are in the care that was being delivered.

And then also along this issue about the waiting of the cost and quality measures in the model, so you have one cost measure or five quality measures or, you know, multiple cost measures and only one quality measure or even how those measures are waited in the outcome the score for that particular model. All matter in how you may be profiling or displaying the result of a performance for that particular model.

Next slide.

So, some of the operational recommendations that came out of that were around recommending that we begin to request the developers, include information particularly on their cost, measure submission for evaluation that they give some more detail around what quality measures are related to that measure either that they have been linking in practice or that may not necessarily be used but then they be related, whether or not they are endorsed. And the recommendation was that this information is evaluated or considered

as a part of usability and used and to better understand the intended use of the cost measure and how it will be used in conjunction with the quality construct.

There was also recommendations around potentially using the measure applications partnership or the MAP to kind of facilitate this discussion and how cost and quality measures are link at the programmatic level. And so – so the previous recommendation obviously is probably more to advance the linking at the measures level and then perhaps the MAP – using the MAP as an avenue to advance this link would be may be – would be more at the programmatic level.

And then the final recommendation which may be a little further off was around creating a pathway for endorsing this composite efficiency measure or some measure that includes both cost and quality measures in a single contract. And so, that potentially, you know, would require criteria and other additional infrastructure to facilitate that. But that – those were kind of the three, I would say, overarching recommendations that came out of that work.

Next slide.

So, I think – so for bringing this issue back to the committee, I think we just wanted to circle back and see where the committee thoughts were on how we should be advancing this work within or through the cost and resource use measure evaluation process.

Do we think that we should try to continue to push this through at least beginning to ask for this information as a part of the cost measure submission form? Should this be something that we try to advance to the programmatic level through the MAP? You know, is there something we want to try to integrate potentially with upcoming evaluation cycles. You feel some of these recommendations may be too aspirational that may be were not there yet or something that we can start now to begin to push the field where we are now.

So, I just want to pause there and see if the committee has any thoughts on that issue.

Brent Asplin: Any comments from the committee members? I have quick one but I want to see if anybody else does first.

Cheryl Damberg: This is Cheryl. Sorry, I'm not feeling well today my voice is a little funny. So, I think in some respects, this work is aspirational, however, I think it's a place we all believe we want to get to. And it seems to me that there's a fair amount of research that still needs to be done in the space.

And I guess one concern that I would have, I guess in principle, I don't have a problem with asking measure developers to submit quality measures along side with cost measures but if they haven't done the underlying research to look at the relationship and sort of wondering what value was that.

Brent Asplin: And this is Brent. The comment I would make in addition to Cheryl's comment is that I agree with the sequencing that you outlined here that using MAP to kind of think through the programmatic elements of how these things are used in making recommendations around that. We need to be doing that routinely before we tackle trying to get the specifications at the measure level to both quality and cost.

If you look at some of the – how challenging its become for developers and for committees to get through to endorsement, I just think and so we have the programmatic level addressed that's going to be difficult to wrestle both the quality and research use specifications to the ground in the same measure. Perhaps, it can be proven wrong, but just based on experience here, I think it's going to be challenging. So, I support the (C plan).

Ashlie Wilbon: Any other thoughts? OK. That was a lot shorter than I thought it would be guys. Well, if you have any other thoughts that, you know, come across as you may be digest a little bit more about for the grouper discussion or the linking cost and quality discussion, please feel free to send an e-mail to the team. These are issues we're actively thinking about. I think, particularly, it sounds like the leaning of the committee right now is to really try to push the linking cost and quality work through the MAP at the programmatic level and we'll start to think about how we can more actively engage in doing that.

But if you have other thoughts that come up over the next few weeks, please feel free to reach out. And like I said, we will be circling back to you particularly on the episode grouper work as we are thinking (about) how to operationalize the rollout of the evaluation of those measures. So, thank you everyone for your time. I don't know if there's any other final thoughts.

Kim Spalding Bush: This is Kim from CMS. I just had a process question about what is the next steps for the developers to update their responsive and get those back to you after today's discussion. There were just like one or two tweaks that came up.

Ashlie Wilbon: Well, I'm going to let the project team discuss that with you.

Hiral Dudhwala: Hi. This is Hiral at NQF. Kim, if you just want to e-mail your edits to the efficiency mailbox, that would be wonderful and, you know, we can give you any additional guidance if need be.

Kim Spalding Bush: OK. Thank you.

Hiral Dudhwala: Sure.

Cheryl Damberg: And I just want to say thank you to the NQF team for digesting a lot of material and making it really easy for the committee to follow. We really appreciate all the work you guys do.

Ashlie Wilbon: Thank you, Cheryl.

Male: Yes, I agree, Cheryl, thank you. We will appreciate it. Thank you.

Male: Absolutely.

(Crosstalk)

Male: Thanks very much.

Female: Thank you and thank you also to our co-chair for leading all these meetings and in-persons and, you know, that was – we really appreciated that Brent and Cheryl.

I guess we could just briefly transition into next steps before we let everyone go. Just so you know what will be happening after this call. So this is our last committee meeting call for this project phase. So after this our team, our NQF team will go ahead and, you know, make updates to the report based on obviously the discussion and we will be getting ready to send the report for our member voting which will take place from June 28th to July 5th. And from there, it will undergo CSAC review on July 11th to 12th. The time is still not known but it will happen during one of those days.

From that, there will be an appeals period which will occur from July 14th to August 14th. And so, the final report will be finalized by September 26th of this year. So those really are the next steps in the timeline.

Hiral Dudhwala: So does anyone have any questions? I did want to also just open it up for public comment after the grouper discussion in linking cost and quality that Ashlie had in case anyone in the public also had any question. So, operator can you open up the line for public comments?

Operator: Again at this time, if you would like to make a public comment, please press star then the number one on your telephone keypad. Again, that's star one to make a public comment.

Suzanne Theberge: And we do have one comment that came in via the chat box. So while the folks are queuing up on the phone, I'll just read that comment off. Someone commented that, I think it is quite complicated and we need to see the proposed methods and standards. So just a brief comment but I wanted to put that out there.

Ashlie Wilbon: So that in relation – which discussion was that in relationship to Suzanne?

Suzanne Theberge: I'm not 100 percent sure. I just see it in the chat box if the commenter wanted to submit an additional comment, we can clarify. It came in about five minutes ago. So, at some point, in the last presentation.

Hiral Dudhwala: OK. Are there any other comments or questions? All right. OK. Well, that comes to the end of our call. So, you know, if anyone has any additional

comments or anything to share with us, feel free to e-mail the efficiency mailbox. And again, thank you everyone for your time and all your participation and, you know, our team really appreciates it. So, thank you all.

Male: Thank you very much. Talk to you soon.

Male: Thank you.

Female: Thank you.

Female: Bye-bye.

Female: Bye.

END