NATIONAL QUALITY FORUM

+ + + + +

COST AND RESOURCE USE PHASE 3 PULMONARY MEASURES STEERING COMMITTEE MEETING

+ + + + +

THURSDAY
JUNE 26, 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 (via teleconference)
JENNIFER EAMES-HUFF, MPH, Pacific Business Group (via teleconference)
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
JANIS ORLOWSKI, MD, MACP Association of American Medical Colleges

CAROLYN PARE, Minnesota Health Action Group
JOHN RATLIFF, MD, FACS, FAANS, American Association of Neurological Surgeons
ANDREW RYAN, PhD, Weill Cornell Medical College
JOE STEPHANSKY, PhD, Michigan Hospital Association
LINA WALKER, PhD, AARP - Public Policy Institute

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

TABLE OF CONTENTS

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

(9:09 a.m.)

MS. TIGHE: Good morning, everyone. This is Lindsey from NQF. We are going to go ahead and begin our Day 2 of our Cost and Resource Use Phase 3 in-person meeting. I will turn it over to Brent and Lisa just to give a brief recap of where we have been and where we are going today, then we can just get started with the Discussion Guide.

CO-CHAIR ASPLIN: Very good. Thank you, Lindsey, I appreciate it. I appreciate everyone's contributions yesterday so that we were able to move through the three measures that were up for our review. And for those of you who did not have a chance to vote yesterday, the link has been sent out and we are doing our voting a little asynchronously with an effort to get everybody's contributions so that we have a quorum. So, if you have not had a chance to complete that, please do so. And then we will have follow-up calls to determine

kind of where the final disposition of those measures is, relative to the voting on the criteria.

Fortunately, we were also able to spend a fair amount of time discussing the longer range view of cost and resource use measures, where we are, where we want to go, what categories of measures are missing. And given the fact that between the period after lunch and afternoon we are able to spend well over two hours on that topic, we do plan to end a little early today. Our goal is to be done by lunch.

And the morning will be spent focusing not only on the follow-up from our conversations from yesterday but getting the additional insight from the Linking Cost and Quality Report that has been drafted.

So, Lisa, do you have other comments to add to that this morning?

CO-CHAIR LATTS: No, I want to agree. I think that was terrific and wanted to

just reiterate that this is our Phase 3 meeting, obviously, and as of right now, there is no planned Phase 4 without the addition of funding. So, I am assuming that NQF is actively pursuing funding for a Phase 4 but, until that happens, right now this Phase 3 is the end of our current standing committeeness.

Can we just check and see who is on the phone?

MEMBER DAMBERG: Cheryl Damberg is on the phone.

CO-CHAIR LATTS: Thank you, Cheryl, and you have been an absolute trooper, so I appreciate you sitting through two long days of meetings, virtually. So, thank you very much for that.

With that, I will turn it over to Taroon. Are you ready?

MR. AMIN: Sure. I will get started with a few thoughts. And if we can, go to the Discussion Guide on page 4, Ann.

So, we had a very robust discussion

yesterday in-between discussions of the measures. And I wanted just again, this is all part of the overall strategic conversation of where we are going with cost and resource use measurement and ensuring that we have a consistent sort of vision of what we can define as success, as we look back to where we are today in five years. And that is essentially the big question we want to ask the group. And we started this conversation yesterday related to reviewing all of the prior work that this committee, members of this committee have participated on and the various other committees, and the various other activities that we have pursued, related to cost and resource use, including the episode grouper activities and the two Robert Wood Johnson-funded activities.

So, we had a robust discussion around our conceptual framework, if you could continue on, it is at the top of page five, the diagram. Thank you.

So, our Conceptual Model that we have been working under and sort of the narrative that we have been thinking here at NQF is around ensuring that we have scientifically sound reliable measures of cost and resource use, essentially, in the green bucket and what is the portfolio that we would ideally like to create. How do we start thinking about the types of measures that we need. During our conversation yesterday, it was clear that we need to continue to have measures of total expenditures and we need to continue to have episodic-type measures that span the high-impact conditions, both in the Medicare program and from those in the private sector. And ideally, those measures would be harmonized to the extent possible so we can assess cost performance across the two different patient populations.

And one of the major take-aways from yesterday's conversation and I think Cheryl was really advocating on this point, was that there

is still quite a bit of work to do in relying solidifying the reliable and valid measures of resource use and really understanding that bucket in itself.

As we look toward the future, one of the things that this committee has emphasized very strongly and NQF stakeholders have emphasized very strongly is that cost and resource use measures must be evaluated and used in the context of quality performance in evaluating the performance of providers.

And one of the outstanding questions that NQF and also the committees have struggled with is exactly what expectation should we be created for developers as they submit measures into the NQF endorsement process. So, more specifically, as developers submit measures of cost and resource use, should they be creating or submitting their methodology for how that cost and resource use measure should be linked to a quality measure or to a series of quality measures, or is this

really a function of how these measures are used? I.e., should these measures be used in a -- should the question of how cost and quality measures be linked be much more of a programmatic question? Meaning that there is some way to link these in terms of how they are reported in the use of the programs.

And that was really an outstanding question. And I think it was a question that was a more forward-looking question that this committee encouraged us to explore during our last round.

Subsequent to that round, NQF staff, along with some support by committee members submitted a proposal to the Robert Wood Johnson Foundation, who has graciously supported the commissioning of a white paper with Andy Ryan and Chris Tompkins. Andy Ryan is one of our committee members and Chris Tompkins from Brandeis, who helped to write a paper looking at an environmental scan of the various methodologies to link cost and quality

measures and to make some recommendations around operational guidance for our go-forward, meaning whether this should be something that we should look at in terms of our endorsement process or something that we should look at in terms of the measures application partnership process.

And many of you around the table participated in that in-person meeting to evaluate the preliminary recommendations by the authors. And we wanted to dive into a discussion, a more detailed discussion around Andy's findings and the committee's deliberations around what is the go-forward strategy related to how we can start linking cost and quality. And I want to just maybe point out the specific questions that want the committee to think about as Andy is setting up and getting ready to just give some overview of the recommendations.

But ultimately, so you will find this on the top of page -- this is the top of

page 8, Ann, you will note that there are a number of questions that we would like the committee to be discussing subsequent to the introduction that Andy will provide related to the operational guidance for this committee.

And I will just walk through them, in particular, to note that now that we have a portfolio of cost and resource use noting the gaps that the committee discussed yesterday and some of the opportunities for additional cost and resource use measures, and that given that resource use measures were conceptualized to be building blocks for efficiency, where do we really want to be in five years? Is building blocks toward efficiency still the goal and the direction that we should be moving forward?

What should NQF be endorsing in terms of efficiency quote/unquote measures, the programmatic methodology of the approach, the combination of measures, or both, or other options?

How might the current endorsement

process for cost and resource use integrate the linking of quality measures? Should the resource use measures be submitted with identified quality measures that would be linked or evaluate cost and quality measures together or continue to evaluate them separately? And this does have also some operational implications for how we might think about the cost and resource use standing committee.

And what are the next steps toward advancing cost and resource use measurement and/or efficiency measurement? And I know that is sort of a broad question. But I would, again, reflect on some of the conversations that we had yesterday, related to what are some of the methodological questions that are still outstanding related to how do we take cost signals and quality signals broadly and really be able to understand provider performance in a reliable and valid way?

So, right now, the methodologies

that we have are limited to selecting individual cost measures and individual quality measures, which is the endorsement process, evaluating them from the four criteria that you have been evaluating yesterday.

And for those of you that may not be aware, NQF also has this Measures Applications Partnership, whose function is to select individual measures for programs and to think about measure sets. However, neither of these functions currently look at the methodology for how individual measures contribute to an overall signal for provider performance for a particular program. And so, the general question here is really when we think about the question of what are the next steps, what are the -- while we may not have to have answers here, what are the outstanding questions that may benefit from additional exploration from additional methodological work?

And so, I would open it up to questions but I would also ask Andy to provide

high-level overview of the work that he and Chris have had the opportunity to undertake and others from the panel who have had opportunity to think about this question, along with the authors and then we can dive right into these questions.

CO-CHAIR ASPLIN: Andy, do you want to go ahead and give us an overview of the report? Thanks for being here.

DR. RYAN: Sure. I have some slides to help us through the process.

So, before I talk about the content, I will just talk briefly about the process. So, as Taroon said, NQF commissioned this paper. The motivation is pretty obvious that the paradigm now in healthcare, it is not just about quality improvement, it is about improving quality in the context of cost, of bending the cost curve. And so to do this, we need measures of cost. We need measure of quality, which is what NQF is now doing. But then really the idea of how you put these two

signals together is, I think the methods to do this are very unclear and what some of the tradeoffs might be in using different methods to do this.

And so, in this process, this was really our goal to get our hands around what are private insurers doing, what are public insurers doing, what are other program sponsors doing to link quality and cost measures to get this notion of efficiency. And then, how can we think about the tradeoffs of these alternative approaches? And then again, what does it mean for NQF and the endorsement process and how NQF might want to approach having some policies about connecting quality and cost measures?

So, NQF asked us to do this and Chris and I developed an outline, we iterated it with our expert panel, which is chaired by Carol Flamm and Joyce DuBow and has some other members. And the Standing Committee are also part of that. I don't know if that is a panel

or a committee or what.

And so, we developed an outline. We got some great feedback. We wrote a first draft of the paper. And then about a month and a half ago, we call came into this room and tried to kind of hash out some of the issues. And we have had, I think, a really excellent discussion. We had numerous suggestions for how we might improve on what we have done and extend what we had done and be more clear.

And after that, we iterated another paper. I want to say that it was also, this was in tandem with Taroon and Ashlie and Erin and Vy at NQF really playing a crucial role in giving you substantive feedback but also facilitating the process.

And then a couple, I guess at the beginning of this month, we submitted another version of the paper, which now is in a public comment period. And so, you know, we may get additional comments there and then there are some ideas from the committee that we still need

to integrate. But what I want to do now is just kind of talk about what is in the report that kind of outlines, say what we did and kind of what we find, as Taroon mentioned, kind of to set the stage for the discussion this morning.

So, can you go to the next slide?

So, the paper consists of a number of sections. Just again, the purpose is kind of what we talked about. So, these key definitions. So, the idea here is that we are using cost and quality to measure efficiency but it is not efficiency from the perspective of production. It is really thinking about the output and efficiency from the perspective of the purchaser. And so, you know, this is the kind of thing, a lot of times these definitions are what ends up taking a lot of time and energy for a committee to kind of manage and get their heads around to kind of what we are talking about and being comfortable with this.

So, a lot of the efficiency literature in healthcare was really kind of how

to produce a unit of output for some given unit level of input, whether that be physician labor, or nurse labor, or even cost. But really what we are talking about in this is producing levels of quality for levels of cost that are actually paid for by someone, whether it be a purchaser or a patient. So, that is really how we are thinking about efficiency in this.

And then the next step is this notion of value that we are working with in the NQF framework where the different stakeholders can take these signals from quality cost and perhaps even efficiency and then kind of combine them to make some preference-weighted assessment of value. So, that is kind of the next step from efficiency.

Again, I think a lot of people think of what we did in this paper as value. And so, kind of just definitionally, kind of reorienting towards saying this is efficiency has been part of the work or part of what we try

to kind of get across in this process.

So, then we have some setups really about kind of the policy context and getting at some of kind of what Taroon indicated about the kind of programmatic imperatives to combine quality and cost measures in our current health system.

And then this section 2 is really about what we did for our environmental scan. So, we did a search both of the published literature, so in PubMed and also in the Gray literature. We actually got a lot of good suggestions from our expert committee about different program sponsors who have been mostly private insurers, who have combined quality and cost measures to proficiency, typically in the context of trying to develop tiered networks.

So, anyway, we pulled together all the examples we could get from various sources and identified what we saw as the mutually exclusive approaches out there to combine quality and cost measures to get this notion of

efficiency.

And so there were a couple of methods that hadn't been used by program sponsors but had just been proposed by academics. And so we just kind of put those in there as well for the sake of comparison.

Then, we illustrate some of these methods. I am going take you through this in a minute. We kind of summarized what we found.

And then I think the important thing that came out at the meeting with the expert panel and our prior conversations with NQF, is to really think about what are the implications for different ways of combining quality and cost measures for these different purposes, so, these different use cases. So, use cases we are thinking were public quality reporting, insurance design, pay for performance, and internal efficiency improvement, and kind of how the approach to combining these measures matter for these different use cases, and, then again, to think about the endorsement process.

So, I am going to take you through each of these sections just briefly in the next few minutes. Can we go to the next slide, please?

So we identify these seven approaches. And I will describe them briefly and then I will show some examples that we worked up on some real data.

So, the Conditional Model is basically, this is, I think, conceptually, very consistent with how kind of NQF is thinking about efficiency where the idea is that cost is assessed for a given level of quality and that is how, kind of efficiency is determined.

So, first, there is a profiling of quality and providers are kind of grouped into certain, let's just say high, medium, and low. And, within these groupings, cost is assessed and through that joint combination of quality and cost, that efficiency is determined.

Now, something that I think remains uncertain with this kind of approach is kind of

once you kind of put providers in these different buckets, how do you then develop a score or you need to kind of go to the next level of saying if you are high cost/high quality, that means something for payment, or if you are high cost/low quality, that might mean something else.

So, but anyway, it is an approach that treats the two dimensions separately and then it is that joint combination used to think about efficiency. Yes?

DR. ORLOWSKI: Andy, the question that I have is doesn't the value that you choose for healthcare depend on individual circumstance? And I will give you two examples.

If my intention is to get a flu vaccine, and I am using simple, my intention is to get a flu vaccine, although I certainly want it to be a quality, you know, I want to have it done sterile, stuff like that, the number one thing is I want efficiency. I want to be in,

I want to be out, and I want the cost of it to be reasonable.

If I have just been diagnosed with cancer, my view of how quickly things should be done and the quality change. Meaning, that I likely want a much higher emphasis on quality and my thoughts about the cost of it may be tempered, although I will still be concerned about it in efficiency.

So, by giving these two examples, what I am saying is I don't know that we will ever be able to vote and have everyone decide what the right model is for value. I think what we have to do is understand that there are a couple of models for it and see if we can apply them to circumstances that may have low moderate, or high medical consequences associated with them. And I think that customers do that when they make decisions about car seats for their children or cars for themselves or whatever. So, I don't know that I can always say what Jim's value, that we are

going to be the same.

I think what we have to say is this is the value of this situation and then consumers, whoever they are, if that is the government paying for it, if that is the public taxpayers, then make a decision where they are going to put their chip on that.

And so as you look at this, how does that come into assessment?

DR. RYAN: That is a great point. And so, I think up to now I haven't been terribly clear about this point but when we are talking about efficiency, we are really limiting this assessment, generally, to a pretty narrow set of care circumstances. So, we are talking about a lot of this measurement is taking place in the context of say, episodes. So, we are talking about comparing providers for an episode of hip replacements, an episode of an acute cardiac event.

So, a lot of the efficiency, the main focus of this, at least, is to think within

a specified set of services how can we assess the relative kind of cost and quality of a set of providers as opposed to over the scope of different healthcare services. How do we think about kind of the relative costs and benefits of doing those.

So, I think that is something that has come up and how do we distinguish this from kind of comparative effectiveness research and kind of the broader issues in thinking about kind of what we should be doing in healthcare and how we should be doing it. And I think the way this effort has been defined has been to kind of look within kind of a set of clinical services that is relatively meaningful definable and then compare providers within those services.

DR. ORLOWSKI: Maybe my examples were too simplistic. So, let's use the example of hip.

There are many reasons for someone to have a hip replacement. Some is mobility.

Some is merely pain relief. And so what I would say is that there are a whole host of individuals who approach a hip replacement with dramatically different expectations for outcome and therefore, value.

There are people who get hip replacement who want to be able to continue their five-mile running a day. There are people who want a hip replacement just so that they can get up and go to the bathroom and do it without pain.

So, even within the narrow confines, using your words, I would still argue the fact that people bring different expectations and, therefore, value is different.

CO-CHAIR LATTS: So, my inclination would be to let you get through your slides, Andy, and sort of present the models and then have discussion. Would you want discussion now, or do you want it as we go along, or do you want to get through the models? What

do you think?

DR. RYAN: I am in no hurry.

CO-CHAIR LATTS: Okay. Okay, then Larry, go ahead.

MEMBER BECKER: So, I want to like understand what you are talking about here. You know I see these things as framework papers sort of direction developers, directions to the community about how we want to work this forward. And it seems to me, along the lines that Janis was talking, is that at the measure developer level, these models, like I read through the document, these models are fine, except that it would seem to me that you would want to engage at the very beginning of what questions is this model going to answer for a condition, both the patient and care teams. Real patients, real care teams, what measures are going to help patients make better decisions and what measures are going to help care teams make better decisions.

And I think, to Janis' point, I

think it is going to be different, depending on who you are in that scenario and what your goals are. And I know that is hard because we did it PCORI and we said you have got to engage patients and the questions have got to be answered and we drove the research community crazy. And maybe we will drive the developer community crazy. But, if you think about it from the other perspective, when all of this is done and we have a series of measures, the usability of those measures for the audiences that need to use them, I think, will be clearer and more effective.

CO-CHAIR LATTS: Do you want to respond to that?

DR. RYAN: So, one of the -- that is an excellent point. And I think something that has been very clear in this process is that when we think about efficiency, we don't want to obscure everything that goes into it and then just say this is efficiency. We have a number. Trust us, this is what it is.

So, what we want to do and I think this has been a very clear recommendation is to not kind of obscure the component parts as you go through the process and get towards a measure of what we say is a reasonable profile of efficiency. And the hope, at least, is that by being transparent, and seeing what is under the hood, and saying okay, so this is the efficiency profile but we can see that their quality, the providers overall quality may have been this. The quality on this dimension may have been this. The quality on this dimension may have been that. We can see cost overall maybe broken out that it would, that there would be more information there than just a single summary score and that information would, hopefully, be useful to numerous stakeholders, potentially for numerous reasons. That is, at least, the objective. And I think that is something that has come out of these processes is kind of what is -- to move towards that objective kind of how do these different

approaches kind of play into that and how can we talk about efficiency without -- in a comprehensive way as possible but at the same time not having kind of too much information and have like just overload data on every measure that just ends up not being useful to anyone.

So, I think those are some of the discussions that we have had and that is something we are trying to grapple with in the paper.

CO-CHAIR LATTS: Lina?

MS. WALKER: Yes, so I want to refer back to the diagram on page 5 of the Discussion Guide. Because it seems to me that Janis' discussion and what Larry also said, you had asked earlier -- yesterday we talked a little bit about it and, Taroon, you mentioned it again, you asked us whether that diagram is an appropriate way to think about efficiency value. And under the efficiency category, it was just quickly, time, quality, cost, to get to the value proposition you include consumer

preferences.

And so it sounds to me like Janis and Larry have trouble with this efficiency concept and they are questioning whether that is an appropriate way to describe value.

But my question really is for Andy because my understanding of what they were asked to do is to look at the narrow concept of efficiency without talking about consumer preferences. Is that right?

DR. RYAN: Well, that is right, but I think in the context of thinking about efficiency for these difference use cases, at least on some level, we were asked to reflect about kind of what consumers, or at least what we are looking for, say, in the context of say public reporting where information about efficiency paired with the quality and cost information is going to be something that people would potentially use to make decisions.

So, I think the call to the paper, again, was not to directly think about the

minutia of consumer decision-making but at least from the perspective of the use case, and how this will actually be implemented, think about which type of profiling method and kind of display of information will work for patients actually trying to make decisions.

CO-CHAIR LATTS: So, my placard was actually up before it fell on the ground. I wanted to make a comment as well.

Because I am not sure, you know similar to what Andy just said, I don't know -- I understand that you want to get a consumer perspective. But I think when you are talking about efficiency, it is really about who is paying for the services because if you are not paying for the services, efficiency means something totally different to you.

Well, are the taxpayers paying? I mean everybody is paying. But what you are paying is very, if you are paying your premium and then you are paying to go see the doctor, most consumers don't connect those two. Most

consumers don't say oh, if I get a more expensive service here and my healthcare costs more money, then my premiums are going to be more next year. People don't make that connection.

MEMBER BECKER: But I don't know anybody that doesn't pay something towards their healthcare --

CO-CHAIR LATTS: Correct.

MEMBER BECKER: -- based on where they are in their financial situation. It means something to them.

CO-CHAIR LATTS: But for the majority of people, they pay the same, whether they are seeing the most expensive provider or the cheapest provider for the majority.

Now, there are people that are in high deductible plans and that is changing the whole conversation. And for them, it means something totally different. But I think that is the point. It is efficiency from the perspective of who is paying. So, if you are

a consumer in a high-deductible plan, you are paying and efficiency matters. If you are a consumer and you are just paying a flat amount, you know, your coinsurance, your copay, I think it matters much less.

So, I think when you are looking at efficiency, your perspective in terms of what you are actually out-of-pocket paying, matters a great deal.

MEMBER BECKER: So, if we are talking about today, that is one thing. But I think for all of this to happen and actually be implemented, we are talking four or five years down the road. And you hear the conversation about the race to the bottom, the 60 percent. So, I think people are going to be paying a huge proportion of their care. And so, they ought to have a metric. It matters. And to make some of these decisions, and I think we need to be looking at it from that perspective.

CO-CHAIR LATTS: And I agree in the sense that I think where they are paying it

matters a great deal. But we know where they are not paying, they often choose the more expensive provider because they use that as a proxy for quality.

MEMBER BECKER: I don't want to get into that argument. I don't agree. So, I will just leave it there.

CO-CHAIR LATTS: Taroon, did you want to comment?

MR. AMIN: Yes, I just wanted to comment on Lina's point. I just wanted to reiterate that one of the big outstanding questions that we were thinking through as part of this original proposal that came out of -- I mean this conceptual framework came out of the first cost and research use effort that we took on in 2010, I believe. And there was a strong sentiment at that time that there was this sort of quality and cost sort of weigh station, which was around efficiency that could be objectively evaluated before you really talk about how you look at stakeholder preferences to

understanding value.

Now, I think, and as part of the conversation with Andy and Chris, we laid out that this was the conceptual framing that the committees have provided input. But, if there was a strong feeling that the community had moved beyond this Conceptual Model, to think about either efficiency or value differently, then we should explore that and challenge our current assumptions.

So, it certainly was the place that we started and we really wanted to examine whether sort of cost and quality could be examined together, and objectively, in order to get toward value, which would introduce the patient preferences that we have been discussing.

So, that was the setup for where Andy and Chris started. And they could talk through how much they agree with this question and what they found, but that is where we started this work.

CO-CHAIR LATTS: All right, Andrea, then Joe, then Ariel.

MEMBER GELZER: Yes, I was just struck by what you said, Larry, about helping consumers make better decisions, helping providers make better decisions. But I think we also have to -- we have to have metrics. We have to have guidance to make systems more efficient and the processes across the board more efficient. And somehow, I think those are still lacking. And we have to -- and you know so that the systems support all this work in efficiency but don't compromise quality.

MEMBER BECKER: Who are the systems?

MEMBER GELZER: Pardon me?

MEMBER BECKER: Who are the systems?

MEMBER GELZER: Who are the systems?

MEMBER BECKER: They are the care teams.

MEMBER GELZER: They are the care teams. They are the care teams. They are the employer groups. They are the regulators. They are the evidence-based guidance we give them. I think it is more than a provider having a discrete performance metric or cost target. I think we have to look at the whole system.

CO-CHAIR LATTS: Joe.

MEMBER STEPHANSKY: Okay, picture this room filled with economists. Yes, with lots of -- well, I think there are serious consideration of putting like metal detectors outside the doors because they knew the kind of battles that would be engaged in if we just let it be a free-for-all.

Essentially, we had to come up with a fairly narrow topic, in order to keep people on topic, so that we wouldn't be going all over. Now, there are pieces of this that I don't agree with. Andy knows that I am a more radical economist than I think either he or Chris are.

And for example, yesterday our

discussion Gene and I had regarding the consumer costs to consume healthcare, we left that out altogether in this project for a purpose because it complicates things so much. We wanted to just narrow it down to where these economists could actually leave with only 40 different opinions involved.

So when you consider what is being written here and what is being presented, give us a little slack as far as how we had to cut it down and we knew that we couldn't accomplish everything. We consider how many academics have been struggling with these issues for years and years and years.

Now, from a practical standpoint yesterday, as we had our CMS and Yale buddies here, they made it very clear, I think, that they want to take the pneumonia cost measure, pneumonia readmissions and pneumonia mortality and combine those in some way to reflect value, at least to Medicare and the politicians that they have to please.

So, it is not five years off that we are going to see some of this stuff come. It may come very quickly. I just don't know in what format.

CO-CHAIR LATTS: Ariel.

MEMBER BAYEWITZ: And it could be there is a committee on this but you know when I look at this, the one thing that isn't clear to me is time. So, there is metrics that could exist around time in terms of evaluating how long does it take to get an appointment, how long does it take to sit in a waiting room, how long does it take to when you meet with your physician for the first time. It probably depends on what your condition is. And when you talk about consumer preference and those different decisions, that is the time that the person wants to evaluate. And are there standardized measures out there around that that NQF has looked at? Because I think that that would be very helpful, especially when making this kind of decision.

CO-CHAIR LATTS: There are questions about it in patient experience measure. Right?

DR. BURSTIN: I think it could be broader than that time to your initial diagnosis of breast cancer after mammogram. I know that there are, I think, lots of potential indicators like that. We don't have any, short of like time to thrombolysis or just PCI for AMI.

CO-CHAIR LATTS: There is time to be seen. Yes, so there is an ER-based time.

DR. ORLOWSKI: Time to first appointment. You know, so if you call, how quickly can you be seen? Time to be seen in the emergency room, time in the operating room; there is all kinds of time measures that can be standardly used across the system.

MEMBER BAYEWITZ: So, are they being used and do those -- are there standards that exist?

DR. ORLOWSKI: I would tell you

that the majority of those time standards are being used by an individual health system or hospital clinic to improve throughput. The only time, the only place that I have seen it used across different health industries has been some insurers have taken a look at time to first appointment with primary care or the emergency room is fairly standard. I would say most of them are currently used internally.

CO-CHAIR LATTS: Brent.

CO-CHAIR ASPLIN: Interesting dialogue. I would say that more contrast purchasing decisions in healthcare with purchasing decisions in other parts of the economy. And I would argue that the models are informative of helping kind of empirically link quality and resource use but really should only be necessary in the context of where we have evidence that purchasers aren't using both the quality and the resource use sides of the equation in their value determinations in their purchasing decisions.

So, let's contrast it to like buying a car. There is objective evidence about quality. There is obvious clear price for whatever car we are going to buy. There is efficiency measures. There is safety data that are out there. And we are informed purchasers or we are saying we are informed enough to make these decisions. There are tons of different value props that are thrown at us, combining those different data elements in different ways and we just make our choices.

I think the better we get at transparency on quality and the better we get at transparency on resource use and cost, the less the empiric connections of the two are going to be needed.

Now, we have a long list of market failures in healthcare and we know there is a rich history of that. And there is current evidence that, with the high deductible plans and consumers that out-of-pocket costs really dominate that equation. So, there may be a use

in that context for thinking through how these models may inform the relationships with cost and quality to consumers in new ways that could be helpful. Yet, if there is evidence there are sophisticated purchaser today, large employers, maybe plans when deciding which network they want to purchase for a particular product that are weighing both sides of this. As long as there is transparency and quality of transparency and cost, the empiric relationship between the two is probably not needed, as we move forward.

I don't know if that resonates, Andy, or not.

DR. RYAN: No, it absolutely does. I think that is a great point. And one of the things that we did, we looked at some examples in other industries. So, we looked at consumer reports and their car ratings and also U.S. News and World Report, their university, they have a value ratings.

And so, it is funny, so both of them

do similar, approximately what you said. But so, there are different quality rankings that are good. They have been validated, they have been used. And then there is the cost side of it and it is compared side-by-side. But in both of these cases and so I agree that having that information side-by-side and being transparent about it, that is -- you are most of the way there, at that point. And we go through some examples about how that information is displayed.

But for both of those efforts to kind of profile efficiency or value, as they call it, they do roll up a value score and then come up with like value rankings, typically. So, with Consumer Reports, it is within certain classes of cars, so, with sedans, SUVs, minivans, whatever.

And so I think it is funny, I agree with you that getting to quality and cost and good measures by themselves, you are almost there. But at least, if we just kind of assume

that a group like consumer reports really has their finger on the pulse of what people really want to make decisions, they have gone that extra step to come up with a roll-up measure that they can then rank, you know, cars on, which further facilitates consumer choice.

So, I absolutely agree with what you said and I think the most important things are getting these measures right. But I think that last step, I think probably remains important and just from a consumer choice perspective, but also from a sponsor perspective, too. Because if CMS is doing something with these measures, CMS needs to make a payment determination, based on quality and cost, for instance, or insurers need to decide what tier a provider goes into based on quality and cost. So, they need to do more than just show the data side-by-side. They have to take the next step.

CO-CHAIR ASPLIN: And so then the unit of the roll-up becomes an interesting question, then, right? And so over the

short-term, if we could get consumers focused in the unit of roll-up relative to bundles and keep purchasers, large purchasers, employers, insurers, et cetera, thinking about total per capita cost over a longer period, over a year in trend year over year, with the hope that in the intermediate term consumers also will be connected to that. We obviously are, in terms of our out-of-pocket cost sharing. But that would be interesting.

I just think getting the heads around the total per capita cost may be just a little too much for the average consumer because they are probably making most of their decisions in real-time about where am I going to get this imaging study done, where am I going to get my colonoscopy. I don't know.

DR. RYAN: Another thing about the use case is that a patient is going to care more about their out-of-pocket and that may or may not correspond directly to a measure of per capita cost.

CO-CHAIR ASPLIN: It is premium and deductible, right? That is what they are making their annual decision on if it is an individual.

DR. RYAN: But one thing that is of practical consideration in this is around cost standardization and if a lot of these measures, these approaches that combine quality and cost measures will standardize cost and then we are basically talking about different resource use for providers, and that is clearly -- that is less relevant for patients who actually have to bear those different prices. So, that is, again, thinking about what is the most useful way to show this information and combine it for the different audiences, different use cases is kind of what we are trying to get through here.

CO-CHAIR LATTS: Gene, then Carolyn.

MEMBER NELSON: Yesterday we started this discussion and we hadn't read the white paper, most of us. And I have just been

glancing through it and it is really terrific. That is kind of one that I think is really helpful. And I have not read it closely yet but already in reading through it quickly, I have learned a lot.

Yesterday, I was advocating for what you would call side-by-side approach. For some of the reasons that were brought up earlier, different people with health problems or different stakeholders will come at this from a different perspective. And so, I was thinking the side-by-side approach with the right component dimensions of quality and cost to value being the most important thing. And I think you said that just now and I certainly agree with that. And to some extent, it is different tools or different data displays for different people with different jobs to do.

And so, to over emphasize a side-by-side versus a rolled up approach, it may be that it depends on where you in the system as to what you need. And getting the subparts

right, then is extremely important.

Just the last comment, that green box, resource use cost/resources used to provide care. That green box, when I first looked at it carefully, resources used to provide care, I do have some friends that are health economists and they would say that is all that costs are, the resources used to provide care in a classic sense. And from the NQF's point of view, and I think from your reports point of view, you are really saying no, in that green box, it is from the purchaser's perspective. It is my out-of-pockets, it is my insurer that is behind my out-of-pockets. It is the purchaser. It is close to the end user, if the beneficiary is the end user.

So, that green box really comes from a good perspective. In a classic economic perspective, it is costs are the resources required to produce the service. But in this context, we are really closer to what does it cost the patient, the family, the insurer

behind them? And so that probably needs fixing is what I think I am sensing, I think, where your report is coming from.

CO-CHAIR LATTS: Any suggestions for how -- what that should be, instead of --

MEMBER NELSON: I would probably defer to Andy. I tend to think about it as the expenditures by the patient or by the family or by the insurer behind the patient or the family.

CO-CHAIR ASPLIN: But basically, you are saying the green box right now is showing how much it costs forward to create the car. And you are saying we need to know how much it costs a consumer to buy the car.

MEMBER NELSON: Exactly right. Exactly.

CO-CHAIR LATTS: All right, good points. Carolyn.

MEMBER PARE: I have a lot of thoughts and there have been a lot of comments and so I will try to be concise.

We were having this discussion a

little bit over dinner last night, just talking about a lot of the things that we discuss at this table. And Lisa, I think you said this yesterday, we tend to move into a mode where we want to boil the ocean because value is something that is different. I mean, every stakeholder in the room could probably identify value differently. And so, I do think that we have to have some consensus around our definition of value for the purpose of creating measures to attain that value.

From my perspective, efficiency really doesn't mean a lot to a consumer and even a purchaser, although, there are very sophisticated purchasers out there. And I think CMS has gone a long way in changing their payment models to move to more of a value-driven system, as have some employers.

The reality is, employers still only have a couple different levels they can use to contain cost and those are eligibility and benefit design, copays. So, again, what is

covered, what is not.

At the end of the day, what an employer is really buying on, whether you like this or not is price and access. That is still, those are two that still the things that rule.

So, at this point in the evolution of quality measurement and the definition of value, I wouldn't rush to the employer and the consumer, though I agree, they are the end user. I don't think they are ready for this set of things that we are talking about because we are talking about value from the delivery system's perspective, the things that we understand, the things that are important, the things that are within our locus of control.

If we wanted to, or if NQF wanted to go in the direction of doing really patient-centered kinds of measures like time between getting a mammogram and then getting an intervention, then those should be built based on what the consumer is asking for, not what we think the consumer should be asking for.

Because if you look at patient surveys right now for satisfaction, those we say are based on what the patient is looking for. So, how nice is our lobby and did the receptionist smile at you and things like that. But I think we would all agree those aren't real quality standards but they are important to the patient. And so, there is still that disconnect and I think we all want to bring it closer so that we can all participate in the value equation but I don't think we should let our aspirations get in the way of what we need to accomplish today to move us forward.

And so, there is a lot in there and I am sure that some might disagree because sometimes I feel it is a little sobering because I always look on the aspirational end. But the truth is, we are still at a place where I think we would like to not be.

CO-CHAIR LATTS: Great comments. Janis.

DR. ORLOWSKI: Brent's comments

actually had me thinking of a possible approach to this. So, I am going to go back to the car.

I think that there are some people who ultimately choose a car because they are looking for energy efficiency. Of all the things, that is on their brain. Others choose safety. Others choose price. Others have an idea of luxury or what they look like. There is a personal affinity with a car.

And I think that you can then say if you are looking at a car for energy efficiency, blah, blah, blah, this or that, and there are consumer price guidelines to this and we have all read them.

And one of the things that I think that we have to ask ourselves is what is the intent of NQF in having the discussion about efficiency. Do we want to add our voice to all the voices that are out there? Do we want to, perhaps, provide a guideline for how the discussion is? You know on and on, there could be many reasons for doing this.

And what I would suggest is that if we are going to do a white paper, you do a white paper because you want to set down certain foundations and guidelines that then leads other discussions about it.

So one approach to this is to recognize that there are many drivers to efficiency. And I would call out those drivers and say these are all the drivers that we can think of. You might have other drivers to this. We are going to choose one driver and we are going to evaluate that driver for efficiency, recognizing that if it is an acceptable guideline, that we come back and then use a second driver and a third driver.

And I think by being very up-front, if that is what we want to do about what we are doing, we can avoid the problem that you have 50 people who say this is the driver that I want, so you acknowledge it. You actually come up with a guideline that then what you are doing is you are actually discussing sort of the

underlying statistics or whatever approach. So, rather than having to come back all the time and talk about the different drivers, you have acknowledged that. You say we are coming back to those drivers.

Then, what you end up doing is coming up with an approach or a statistic and then you say, does that make sense? Yes. Is that approach an approach that you can use for the other drivers or do we have to use other approaches? So, that you end up having a white paper that is a true white paper. It can be used for discussion. People see what is going on. And then again, using Carolyn's idea, if some of those drivers are insurers or employee-specific, then you can say what drivers then get added as patients and families become involved. And I think that is a way to say we can't establish world peace in a day, so, let's just take a piece of it, but that we provide a framework for others who are going to use this methodology to come behind. And I

think that is a way to use a guideline, a white paper without having to have world peace be the ultimate conclusion.

CO-CHAIR LATTS: Did you want to comment specifically on that comment?

MR. AMIN: Yes, I just wanted to reflect a bit on, again, kind of where or what was the impetus of this work. During the first cost and resource use endorsement project, there was a significant amount of concern from the NQF stakeholders related to simply endorsing cost and resource use measures in isolation, the argument being that simply looking at costs without understanding quality performance will drive the healthcare system down to the lowest cost provider, which may have significant quality implications to the enterprise that we have built over the last 15 years. And the strong reaction that we got from the membership and then it was reflected by the governing structures of the organization, particularly the Consensus

Standards Approval Committee, and the Board was around NQF if we are going to move down this road, there needs to be a clear pathway for how these measures should be use and how these measures link to our foundation about measuring quality.

So, the basic foundation of this paper was to start to address that question. And, obviously, there is no one paper, no one effort to be -- there is a stepwise to get there. But the basic foundation question here is looking at the various approaches that are in the field, can we start to characterize those approaches and start to make some recommendations about ensuring that we can look at costs because they are important to measure in their own right? Without any question, cost concerns in this country are significant in their own right without any question but ensuring that as we look at the questions of cost, we are ensuring that we are maintaining or improving quality at the same time. And to

characterize the field of how that is done right now because there is many private programs that have been in use much before public programs have been introduced, but obviously, there is a bigger emphasis on public programs currently.

And first characterize that and make some recommendations around the methodological challenges and approaches and recommendations for a path forward around how these concepts can be linked more systematically.

And so, I just wanted to provide again that as a little bit of context for how we have gotten to this point. And as we walk through some of these models that Andy has characterized based on what has been used in the field, that is really the reason why we are here with these questions.

CO-CHAIR LATTS: All right. It was a pretty prescient action, given what has happened over the past couple of years.

Lina, did you still want to comment?

Okay.

All right, do you want to keep going through your slides?

DR. RYAN: All right, so where was I?

So, this notion of the Conditional Model is looking at basically the joint relationship between quality and cost, profiling providers on that or even health systems, some level of accountability on that combination and then coming up with either tiers based on that combination, or a score, or something.

So, the Unconditional Model just uses performance measures on quality or cost and then assigns weights to those and then rolls them up. So, for instance, a provider could have some normalized cost measure that is very low. They could have a normalized quality measure that is very high. And if they both get equal weight, you put them together, they could then have basically an efficiency score that is

approximately average.

So, this has different implications in that the joint relationship between quality and cost no longer has any particular meaning. It is the independent relationship of both domains. So, for instance, the Hospital Value-Based Purchasing program uses this approach, whereas the physician value-based payment modifier uses the Conditional Model approach, just to give some idea of even within what CMS is doing, there is different approaches that are being used.

And then so, this next idea of the Quality Hurdle or Cost Hurdle Model. Again, so the Quality Hurdle Model is what is used in the Medicare Shared Savings and Pioneer ACO programs where basically there is some lower threshold of quality performance below which providers get, if you don't hit that hurdle, then basically you get no credit for your cost performance. And then above that, once you get beyond that hurdle, there is some profiling of

cost. So, that is the quality hurdle. And so, this is common in Shared Savings-type programs where payers are interested in profiling on costs but they want to assure some minimum level of quality.

So, a variation on this is to continue at one tier, and kind of over the hurdle, you can still condition shared savings based on some quality performance or you could just say all you need to do is get to that quality hurdle, then we don't care anymore.

And then so, on the flip side is the cost hurdle approach, which we saw some sponsors use this where you basically just need to get some cost threshold and then after which the kind of efficiency profile is based solely on quality.

Now, the next two approaches were those that haven't actually been used in the literature but were proposed by researchers. And I am not going to talk about the Regression Model but the Cost-Effectiveness Model, I think

is interesting and interesting to reflect on the implications that a lot of times when we were dealing with these measures of quality, it is some kind of score. But you know, what that actually means in terms of how it relates to patient outcomes or things that we care about in the system. I mean, at the time of roll-up, it really might be kind of uncertain and that is just something that we just deal with, you know, these kind of composite measures that are based on at some combination of patient experience, process outcomes, get to some quality score and then we just work with that. And then we can go through all this effort of profiling efficiency.

And the idea with this Cost-Effectiveness Model is try to basically put a weight on quality performance that corresponds to some kind of monetary value from a kind of cost-benefit approach. Again, this, I know, starts to make people uncomfortable when you are weighting qualities and things

like that. But just kind of conceptually, the idea is that there could be a provider that has very high costs, costs that are much higher than other people and say a quality difference on something like mortality that is only slightly higher than other providers. And by standard efficiency measures, they might look bad. They might look like they are low efficiency. But if we correctly valued the incremental benefits in terms of the quality they are providing because they are providing care that leads to lower mortality, then in reality, the benefit they are generating for that, their expenditures that they have, are actually quite efficient. They are really worth it. And so, that is the kind of model that I think is, at least, interesting to reflect on when it really tries to kind of put a more precise value on what quality is, what quality means.

And so then just to finish, there is the Data Envelopment Analysis, there=s Stochastic Frontier Model and the idea here is

that there is, when we have kind of quality -- you can generate quality and cost performance scores. And then there is an empirical relationship between let's just say quality on the x-axis and costs on the y. And there is some kind of frontier, a frontier efficiency. So, at every level of quality, there is kind of an efficient way that that is produced in an empirical distribution of providers or healthcare systems or whatever we are evaluating.

And so the DEA Model is identifying that difference from the frontier and then saying that we are determining efficiency based on the difference from the efficiency frontier.

So, it has been used a lot in these studies of thinking about efficiency from the production cost perspective, where the outputs are something like hospital days or visits or something like that, and the input costs are labor or just expenditures or whatever. And the VA actually does this, uses this approach,

for their kind of efficiency profiling in that context.

Well --- but there aren't any program sponsors that are using this where quality is that kind of dependent variable and costs are the inputs.

And then finally, we have the side-by-side model, in which there isn't an explicit way of combining or jointly profiling the cost and quality domains but they are just shown side-by-side. And a fair amount of programs we found used that and I am sure there are others that use it that we didn't identify in the scan. Can we go to the next slide?

So, we identified 24 programs that are currently doing something to combine quality and cost. So, the complete list is in our paper. These are just some examples, the Blue Cross Massachusetts Alternative Quality Contracts has gotten a lot of attention. You probably know about that. Aetna Aexcel, Anthem, Blue Cross, these are basically ways of

establishing the kind of tiers for patients to face kind of differential cost-sharing on the kind of whether providers are deemed to be efficient or not.

And then towards the bottom, we see Hospital Value-Based Purchasing. So, this is a program that is near and dear to my heart. And in the first year was just hospitals received this total performance score that was based just on clinical process quality and patient experience. And then in the second year, there is an outcome domain that was added. And in the third year, now there is going to be a cost part of it. And as I mentioned before, the way of combining this is the Unconditional Model.

So, hospitals are just scored on these different weights. The cost weight is based on an NQF-endorsed measure of Medicare spending per total Medicare spending --- per something. And it is basically just an episode, price-standardized episode cost,

including the three days prior to admission, 30-days post-admission.

And then again, the Medicare Shared Savings and Pioneer Programs use the Quality Hurdle Model. So, can we go to the next slide?

So, again, something I think was interesting was to see the kind of variation that was out there in the field. I mean, there is not really consensus. There doesn't seem to be much guidance. And, you know, people --- and this is, again, part of the motivating idea behind this is to try to provide some of the principles and ideas to consider. But you know, there is some mix of the Conditional and Unconditional models that were used side-by-side. I think, again, the Quality Hurdle was used a lot in these shared savings programs. That just seemed to make sense.

And then we had a mix. We saw a mix of these programs that profiled physicians, hospitals, some combination of these and health systems and plans.

One thing I want to note that I thought was interesting, there were a number of private programs that were profiling specialties that have basically not been touched in the public programs. So, there is a lot of interesting kind of specialty profiling that was used that --- CMS is stuck pretty closely with just hospital-based profiling for standard conditions. And there was, I think, more creativity, more variation on the private side that I was actually surprised to see. Can we go to the --- yes, please?

MEMBER GARRETT: So, I know like in Value-Based Purchasing it also looks at not just the absolute performance on quality but also at improvement. Did you look at that question across the different models and whether they were looking at absolute or improvement?

DR. RYAN: So, we defined how the two dimensions were specified in general terms.

So, you are right, so Value-based Purchasing uses this model. So, it takes for each individual measure, hospitals get assigned the greater of the achievement points or improvement points. And so, you can do well for either.

There may have been one other program that did that but almost everything else was just levels.

(Off-microphone comment)

Well, I think that profiling on improvement, in addition to achievement like really makes sense conceptually. I mean just what we talked about a lot in the field, if you are just paying on levels, particularly if you are using levels that are limited incentives for lower initial performing providers to get better. And so paying for incremental changes really makes sense. So, there is kind of two ways to do that. You kind of just pay on levels but just have it be, you know, you are paying incrementally more for each level, so there are

still incentives to do better regardless of where you start. But then there is also paying explicitly for improvement.

So, I like it. I think it makes sense. I think that empirically, there is no evidence to show that it actually generates more improvement to do it that way. But you know, I like the idea in theory.

I just wanted to --

CO-CHAIR ASPLIN: Mary Ann.

MEMBER CLARK: I just wanted --

DR. RYAN: Sure.

CO-CHAIR ASPLIN: Why don't we take Mary Ann's question? And then, Ann, my understanding is that you have some operational recommendations at the end of the day. Is that -- or some framework questions. If you could kind of skip to those after Mary Ann's question so that we probably will have some dialogue around that and then there is an additional conversation we need to have this morning before we break. I just want to make sure that

we are not all of a sudden it is noon and we are not getting value to NQF.

So, Mary Ann -- well, don't misinterpret that. We are not meeting the needs that they have asked us to meet this morning. That=s just a better way to phrase it.

MEMBER CLARK: Can you just go back a slide? I just had a question on the next one, I think, on these different models, the 24 programs and the breakdown of these different models.

Did you look at the audience for these different models? I am just wondering whether there is on specific because it seems like there has been a lot of discussion on who is the audience for these measures. And you know, if you are going to incorporate the patient component into it, some of these may be easier to understand by the patient as opposed to some other ones.

And the one I was just thinking of,

for example, the Unconditional Model that CMS is using in Value-Based Purchasing, I could see that pretty easily being used by a patient because it could incorporate their value of the different components. Right now CMS has very specific components, for outcomes, for patient experience, for the costs and for the process measures. But if a patient valued those measures differently, they could actually input different percentages and say oh, well, this provider is the one I want to go to. I could see that working pretty well for a patient. But I just was curious whether you looked at the audience for these different --

DR. RYAN: You know, we didn't classify the programs, based on the audience, per se. I think Hospital Value-Based Purchasing is a good example and CMS would probably say that the audience --

So, in some sense, in its most basic sense, it is just trying to generate different payment adjustments for hospitals based on how

they do. So, you know hospitals are really the target.

But CMS could also say well, this gives a signal to patients that they could use to make decisions but the way the information is displayed now is not in a way that I think patients could use it in the way that you described.

So, I think -- I'm trying to remember how we classified these programs. But you know, some of them -- that is a good suggestion. I should review that.

But it would good to at least have some point about -- is this just trying to tier providers? Is this trying to just generate a payment adjustment? Is this purely for public reporting or to be more explicit about the use case? And in some cases, it might be multiple things together.

CO-CHAIR ASPLIN: You can go ahead and walk through, if you could just kind of give kind of a high level thumbnail on each one and

then get to the last piece, that would be great.

DR. RYAN: Maybe we could just skip through the examples of the different approaches. I was going to say what we did, is what we did is we used the Medicare, the standardized cost measure that has been endorsed by NQF and is now going to be used for Hospital Value-Based Purchasing but it is not, at this time, wasn't currently part of the HVBP payment formula. And we combined that with the total performance score from Hospital Value-Based Purchasing. We normalized both scores, put them in units of their own standard deviation, so plotted them. So, each of these plots is the same, but what we did is we applied the different profiling methods to the different, to kind of this scatter plot of quality and cost and we used this to say well, if you profile it this way, this would be kind of how the profiling would work out. If you did it this way, that would be the way the profiling would work out.

And the expert panel thought that these depictions were kind of useful in thinking more clearly about what it actually meant to profile this way or that way and so just thinking conceptually.

But maybe we can skip to the -- there is a correlation table at the end here. If you keep going --- one more.

And so, I think there was some question when we met that like does it really matter how you -- what model you use to combine the measures. Do they just have the same information and are there different implications in how you would actually generate a score and then kind of say this is efficiency.

So, one thing we did do was we looked at the scores that were generated by these different models, sometimes under somewhat different assumptions and then ran a correlation matrix. And not surprisingly, they are all positively correlated. But you will also see that there is a pretty high degree

of variation in the correlation between these different models. So, some of them are very highly correlated and some are pretty weakly correlated. And so I think, to us, this was illustrative that number one, that we were -- that although in some ways we think all these models are getting at efficiency, they are getting us in the same direction. That is true. But that these different correlations will certainly result in different rankings of providers, depending on how you would do this. So, it turns out it is important and it does matter how we would combine these measures to generate scores. So, that was one of the inferences that we had here.

So, now I just want to move on to some of the high-level -- can we go to the next slide? So, I guess we already talked about this. Can we go ahead, please? Thank you.

And so, then just to think about some of the kind of -- we have -- so what we did we moved from doing a scan to kind of

operationalizing the measures depicting them and then thinking about what we would want to do. What are some of the kind of bigger ideas that we kind of want to get across that the principles that NQF could potentially provide to developers or otherwise integrate into the endorsement process? And one of the things we came up with was something that we came back with before which was that, when measuring a profiling efficiency, we shouldn't be obscuring the component parts and should keep them separate to allow kind of different determinations to be made rather than miss that output of the combination process.

I think another key point was that choice of method to combine the measure should depend on the use case. And that is really, I think, come out in this discussion. And that is, I think, just an important consideration.

And I think one of the implications here is some of the different methods have placed more weight on different dimensions.

So, for instance, the Quality Hurdle placed kind of more weight on costs if we you are just trying to get over that hurdle, over the lower threshold and then beyond that, quality doesn't matter. So, it potentially creates different incentives for providers in the system, the ultimate ways that quality and cost measures are combined.

I think, you know, this was the idea of discrete versus continuous measures of efficiency was something that we talked about a lot. And again, this also really depends on a use case.

If insurers are trying to come up with value tiers, basically they need a classification, they need discrete tiers to say you are in, you are out. There isn't a gradation. But in other cases for, say, payment adjustments or even public reporting, that the arbitrary classifications can -- they obscure information and they can add noise to the levels that you show.

At the same time, I think -- and we talked about, when we talked about the different use cases, it is hard to not talk about the larger issues in program design and this display of information I think remains really crucial. And so, having, especially with these side-by-side methods, and we don't want to just have side-by-side comparison of hundreds of measures, which is kind of how Hospital Compare is currently structured. We are adding, now we are adding even more information; we have quality, we have cost. And particularly if we are trying to get to a consumer decision-making process, this information needs to be displayed in a way that people really understand. I think the stars are actually good, as long as there is some star increments in there and it is not just good or bad and we can get to a star system that people seem to respond to. Can we go to the next slide?

And then finally, I don't want to

say quite in lieu of endorsing efficiency measures, but I think when we started this, it was part of what we were trying to think about and the system was what was out there. And right now, there aren't many measures that are just, that people say are -- this is efficiency, and this is a score and this is the efficiency score. I mean, generally, what we saw was some combination of both of the domains of quality and cost to get to efficiency. So, that might seem like a fine distinction but it is what we saw.

So, NQF could have a couple different routes here. I mean, they could take the approach of just endorsing efficiency measures. But at the same time, that doesn't really seem to be kind of where the field is. The field seems to be keeping these measures separate and then combining them.

So, I think in lieu of an official kind of endorsement process, NQF could do a couple things. I mean, so I see program

sponsors here, but it could also be the developers is to do more than -- let's just keep on the program sponsor's side, is to really go through this process when thinking about combining quality and cost measures and being clear about the use case. The cost and quality measures would presumably be those that were already NQF endorsed or there could be some compelling reason to not use them.

And then there would be this approach to rolling out the domains and then articulating a reason why they would want to use one of the established methods identified that is geared for that use case.

And so, NQF could either kind of recommend this for program sponsors or it could also put more of an onus on developers when presenting cost measures to do more than just say this is our cost measure but more so, this is the process through which we propose that this cost measure could be paired with quality measures to get us towards an efficiency

signal.

So, that is kind of where we are right now in the report and we are looking forward to a public comment and more input from the committee. But this is the kind of where the process has led us so far.

CO-CHAIR ASPLIN: Andy, thank you very much. I appreciate you walking through that. I think this a very helpful framework here at the end.

I am going to suggest not for the purposes of the paper, because I think anybody who can hang with your paper can hang with the efficiency definition fine. But I think we have a language challenge when it comes to the word efficiency. And I will just give you the past two decades of trying to deal use of the emergency department for non-urgent conditions and the efficiency arguments all over that because I just lived it year after year, after year, after year.

Actually, I mean if you have an open

emergency department and you are asking a question what is the cost of producing care at the margin for non-urgent conditions, it is incredibly efficient, incredibly efficient to see those minor conditions at the margin, when I already have an up and running emergency department.

That is an entirely different question than is that an efficient purchasing decision for those conditions. Right? And 20 years, there is nothing but murkiness in this. That distinction is completely lost. And I had given up on trying to -- I mean I keep saying to the emergency medicine community, look, folks, it doesn't matter how much it costs you to produce the care. It is not an efficient purchasing decision, based on the business model you are in, in many cases.

So, I get the argument. I just -- can you walk me through one more time, and I think it was the consumer preferences and so forth, and the kind of the clean analysis of

efficiency from a purchasing perspective as opposed to production perspective? But why the external view couldn't be like a value framework, one more time just walk me through. Because the efficiency on the production side and value on the purchasing side was, at least from a language framework, seems to be more practical. Maybe not for the paper but just for language that we are going to use. Because otherwise I think everybody is going to apply their definition of efficiency to the conversation and I am afraid we just get really confused.

DR. RYAN: Well you know, Brent, I thought that your previous articulation of efficiency there we're -- the purpose of this paper was to look at it from the expenditure side and not the production side. And so that is really our lens here.

So, on the expenditure side, we are saying for the end user, be that the payer, the patient, what kind of level of quality are we

getting for this outlay. Right?

CO-CHAIR ASPLIN: I'm not asking you -- I think your paper is elegant. Right? I think you have articulated it really pretty clearly. And for purposes of conversations like this, I don't think you need to change a thing. Right?

I am asking the question: how is NQF and how are we, as a community, interested in trying to facilitate purchasing decisions that are based on both dimensions of the value equation? What language should we use to move that forward the fastest?

And all I am saying is that in that broad multi-stakeholder dialogue, the word efficiency has way too many meanings and we have already started to create the definition around value. Do you know what I mean? It is not a comment about the paper.

DR. RYAN: No, I know. What we try to do is just kind of bracket that production sense of efficiency and say like yes, you know,

that is kind of the economic definition of efficiency. And a lot of people understand it that way but that is not how we are using it here.

But I agree. Even after all this process, we had a call the other day. And someone on the committee was like you know what are you saying is efficiency. You know I still don't really think that is efficiency. I think it has value.

So, you know, we are trying --

CO-CHAIR ASPLIN: I get it. I think I get it.

Gene, do you have a -- what are your thoughts?

MEMBER NELSON: Yes, I have a comment. I think to your point, we could, and I think maybe should, drop the term efficiency. I say that a little bit reluctantly because of the IOM quality definition, safe, timely, effective, efficient, make it all patient-centered. But I think we should be

focusing on value and use that as the term. And that is what I think, in general, society is asking for from us providers.

And so to focus primarily on value and to move out the term of efficiency, a little bit more on that model what Taroon was saying, how should we think about it? And again, I really like the IOM's framing of the dimensions or aims at quality, state, but we now have a national quality strategy that more or less is the three-part aim or the Triple Aim. And so we might conceptually describe value as a function of outcomes and experience in carefully thinking about what we -- all that goes into experience and costs, meaning expenditures. And that that is probably more with the contemporary times to be focusing on enabling measures of value in line with the three-part aim.

DR. RYAN: I just thought of one other point I should have made and it is this notion of objective measurement. And I think

for a number of people that are part of the stakeholder process, the idea of objective measurement of value is just kind of problematic, whereas the idea of objective measurement of efficiency feels more okay.

And I think that is kind of one of the reasons why this is what it is. Value is something that can never really -- it is like everyone has their own idea of it. So, we can never specify precisely what it is, whereas, we potentially could do that for efficiency because this I trying to get us towards something quantitative. That is one of the reasons why that efficiency I used.

MEMBER NELSON: Well, that is why I used the term it is a function of. In the function of outcomes experience cost depends on your perspective and your circumstances. And that is where the waiting comes in. It is a function of. It is not --

DR. RYAN: There is something else I actually wanted to say, there is some

information there is some preferences there.

MEMBER NELSON: Right, there is the preferences of the concerned individual, from their perspective.

CO-CHAIR ASPLIN: It's almost as if, even adding the words purchasing and production -- I'm not an economist, obviously -- but the purchasing efficiencies that we have been talking about, you have been describing, versus production efficiency.

And kind of in the empirical architecture that we are basing all this on, I think those terms are fine. The question is as we translate that to having a dialogue with the public and our consumers, our patients, and our members, our employers, I don't know. I just think there is great risk of confusing folks because we all bring our own definitions of efficiency. For that matter, we bring our own definitions of value. So, I am not suggesting that there is crystal clarity here in any case.

Carolyn, do you have a comment?

MEMBER PARE: I would hate for us to walk away from the term of efficiency. I think yesterday, first of all, Gene, I am glad you said what you said about the Triple Aim, because yesterday, as we were having discussions, I was thinking about the best way to array this for people to understand is within the context of the Triple Aim. So, whatever measures we use, they would fall into a quality cost or a patient-centric parameter. And then people could choose, select their value based on the combination off that matrix.

So, as construct for how we report it, I think it is really useful but I wouldn't want us to do away with the term efficiency in lieu of or in exchange of value because I liked what Taroon said yesterday about efficiency as being a way station on the path to quality. I think efficiency is critically important in understanding and I think where we are having difficulty is because efficiency cuts across two of those Triple Aim parameters, cost and

quality. It is the combination of the two that gets us to efficiency. Efficient utilization of resources is really what we are looking at here.

So, I think we are getting kind of caught up in maybe some of the -- I don't even know what the right words for this is but how we frame this. But I think efficiency is critically important for us to keep an eye on.

CO-CHAIR LATTS: Yes, very.

MEMBER STEPHANSKY: So, HFMA adjusted two papers on transparency. And I thought they did a relatively fabulous job, one aimed at the patient and one aimed at the provider community. And maybe we should think about doing something in the same vein that does lay all of this out in terms that each of those stakeholders can see themselves in.

CO-CHAIR LATTS: Carolyn.

MEMBER PARE: I also keep drawing on my experience with the Buyers Health Care Action Group when they put that model in place,

the first tiered program everywhere. We actually did array providers based on cost and quality. And then, again, within the matrix, you could show whatever quality parameters we had, although the quality was really built into the pricing mechanisms and how the providers bid their price is too long and involved to talk about now. But we did show them in the tiers. And then we had our little quality signals for those that got quality excellence awards. And then of course we showed how other consumers talked about those particular care systems.

It was really a good -- now, obviously, it didn't last for long. Things got changed but it was successful in moving the market into more of these tiered kind of products and things. But the things still that moved the consumer behavior was cost. So, they moved for price. That was probably one of our biggest learnings is that they did always move for price over anything else.

CO-CHAIR LATTS: Cheryl, did you

have a comment?

MEMBER DAMBERG: Yes, thanks.

So, listening to this conversation and Andy, I did read your paper last night, although I have been sick and so I probably didn't catch everything that was said there, that you are a brave soul for wading into this morass.

And I think that is my biggest concern is that this is a morass and that what we are trying to get our hands around is really a space that is so vast and complex. And I am wondering if there is some value in NQF taking three steps back and trying to simplify kind of what is it that is NQF's charge to do in this space?

And a couple of things that sort of have come out of this conversation and my read of the paper is that, in essence, we are trying to construct a set of measures that are trying to drive the marketplace to improve quality and reduce costs. And those measures can be

developed independently or some combined measure that Andy has outlined a whole array that are being used in the marketplace.

But in essence, we are trying to signal to providers the actions we want them to take. And so my question is, can we sort of go back to first principles and outline what are the different uses that we see of cost and quality measures. And then what kinds of measurement properties would we want these measures to have? Because I think that the market is going to try to innovate and use these measures in lots of different ways that we won't be able to control. But I think if we had a set of guidelines or principles that we wanted them to try to adhere to in implementing measures. So, Andy's, I think it was the previous slide, pointed out, one of the things about introducing noise into the signal and the risk of misclassification, let's say if you use star ratings, as opposed to the underlying continuous data. I mean, those are the types

of things that I think NQF is best positioned to signal the market on is what are the appropriate measurement properties. We want the measures that are in use for different applications to possess.

CO-CHAIR LATTS: Thank you. Nancy.

MEMBER GARRETT: So, I just wanted to react to this idea in the paper of NQF recommending a process to follow and I like this idea.

So, in past jobs that I have had, I have been at two private health plans where I have developed the methodology for doing tiering. And it is kind of the Wild West. I mean each plan is making it up and increasing that they are sharing best practices but there isn't really a set of principles and approaches for that.

And then now, as a provider, having each plan take a totally different approach, a provider can be rated completely differently

just because of the difference in the way the measures are applied and the methodology.

And so I think having some guidance, even this question about whether quality should be based on improvement or absolute performance, you know, what are the best practices, and can NQF play a role in starting to get more harmonization and more best practices out there in use? So, I think this is a really good idea. And I don't know the right place to do it, if that is this committee, or if that is a different committee, but that I would support that view.

CO-CHAIR LATTS: Great. Thank you. Gene.

MEMBER NELSON: I had a suggestion for possible measures and principles. Would that be an appropriate time to make that suggestion or not?

In going back to this idea of side-by-side measures, and one principle being that a framework for thinking about value would

be the triple aim. So, that is one principle.

And another has to do with cascading measures, so that we might have a high-level measure that can be cascaded down to lower levels. So, we might have Dartmouth-Hitchcock Health System's per capita cost and we might have the annual cost for asthma care, within Dartmouth-Hitchcock, and we might have lower levels like care for AMIs, 30 days.

So, that I give cascading measures in getting down to individual care team members, from system to care teams, would be an example of cascading measures, and having the measures cover the three-part aim.

So, then there has been a group working for many years called the Gretzky Group, and thereafter and have been thinking about interdisciplinary group of providers, payers, governmental entities, et cetera. And over time, the recommendations were for a set of measures like this. For outcomes, three categories of measures: risk status, disease

status, and functional status, including health-related quality of life.

Three big outcome measures, possibly harm might be part of that to get at a site of safety, harm done by the system, but risk status, disease status, functional status, possibly harm.

Experience, quite selective. Access to care as needed, decision quality, and self-management confidence. Activation, self-management, confidence. Three key experience. And that is not to say many other aspects of experience aren't important. They are. This coordination would probably be added to that list by most people.

Then, under healthcare expenditures, annual per capital healthcare expenditures and episode cost. And then, as mentioned yesterday, indirect social costs, like productivity, losses due to illness. So, that is about ten measures that might be -- or measure domains, that might be very useful and

be aligned with the triple aim and many of them are cascadable, if properly designed and properly fielded.

CO-CHAIR LATTS: Great. Thank you.

MS. WALKER: A follow-up question. And how are they combining the measures? I mean this what the discussion is about.

MEMBER NELSON: Right. The first is to make a value table and don't combine them. It is a side-by-side idea. Enable -- if I am a health system at Dartmouth-Hitchcock, now we have our value table and we can cascade it down. We can go to work on it. We do make our outcomes transparent, at Dartmouth-Hitchcock, as a policy.

So, we will flip the value tables out using the kind of displays that are consumer friendly, and let people sort of see what it looks like for the system or for, if you are going to get a total joint replacement, or if it is maternity care, et cetera, to put the

value table out for different subpopulations and not, generally, to combine them into one rolled up metric.

MR. AMIN: Okay, thank you. All right. Well, thank you all very much for a very interesting and robust discussion around how do we get to the next step in terms of efficiency and really thinking through the language that we are using.

This will obviously be very good input in finalizing the report as one step. But obviously as input to staff as we think about the next steps in terms of evolving the future proposals that we may consider seeking funding for, and also as we think about the measure endorsement process and the measure selection process.

I think with that, I think we have some specific -- I think we can shift topics at this point, I will turn it over to Lindsey to lead the next discussion around submission elements and our criteria. I mean clearly,

this conversation related to our criterion submission element, but as it relates specifically to the resource use measures that we have been endorsing over the last two years.

MS. TIGHE: Yes, absolutely. So, we wanted to take some time today, understanding that we don't have any funded work in cost and resource use coming up in the foreseeable future, to really just understand some of the challenges that we faced in applying the criteria that we currently use to evaluate cost and resource use measures.

And also, really, in this framing of this broader context of looking at efficiency or value, take a little bit of time to, again, understand what kind of submission elements might be necessary to begin to evaluate these measures as we move forward in this work. So, I did get this point, after reviewing three measures yesterday, you all are quite familiar with the resource use measure evaluation criteria.

As it stands, as I think we outlined in some of the challenges, funnel slide yesterday, we do understand that there is some issues with how to apply the validity criteria; how to understand attribution in the context of the measures, given that attribution is usually a programmatic concept, not a measure-level concept; and really the whole use question, how to understand how to evaluate a cost and resource use measure, agnostic of its intended use, given that you all are well aware of the intended use for many of these measures.

So, I won't go into too much background, given that you all have been applying this criteria and it is listed here in the document. I just want to open it up for conversation on those topic areas.

MEMBER McHUGH: So, I think just on the use issue, that was clearly a recurrent theme in our various evaluations. I think relative to today's discussion, it seems like, and what we are often leaning towards, was how

necessary that was to consider, particularly as these are combined with quality measures.

So, you know I think in the questions that Taroon had outlined earlier about next steps for NQF and what to ask of providers around how they would use these things in combination with each other and what would the use case be. I think that would be more satisfying from a review perspective, because we are all kind of thinking about that and would ultimately we would be able to provide better feedback.

MR. AMIN: Can I ask a question related to that comment? So, broadly, we have these four criterion: importance to measure and report, scientific acceptability, feasibility, and usability and use. And this question around use case, so the current NQF guidance is that measures should be evaluated with the idea that we are use case agnostic; that measures that are endorsed should support quality improvement and accountability

applications broadly.

Accountability applications falls under the umbrella of both public reporting and payment applications. So, the way that the current evaluation criteria is constructed is under that principle. Now, that is for quality measures, in addition to cost measures.

And I think in addition to the conversation we were having just prior to this, and I think as we have seen to some of the evaluation of the measures prior, the use case actually does drive the decision about how we weight the criteria in our minds. It is not clear in the evaluation process how the use case is driving the criteria and how we are weighting it.

So, one of the questions that we -- so, let me put out a straw person and just have people react to it. One way people can react to this is to say the reliability and validity of the measure properties needs to be at a higher bar, in order for it to be used for

payment application, as opposed to a public reporting application. That is a little bit of one could maybe argue that that would be the way people are evaluating measures currently, but that is not implicit in the way that the evaluation occurs.

Obviously, that is a very controversial thing to say, but the question is how is -- so, I think the point has been made that the use case does drive the evaluation. The question is, how exactly is it affecting the evaluation? Are we suggesting that the use case would require different criteria? Or is it that the criteria would be weighted differently, depending on the use case? And is there a general agreement about which use case would drive more stringent evaluation of the criteria? I.e., would payment, would a measure being used for payment require a higher bar than others? So, I am trying to understand in a little more granularity the question around use case and how it relates to criteria.

MEMBER DAMBERG: Taroon, this is Cheryl. I don't believe you have that kind of thing on the quality measure side, do you?

MR. AMIN: We don't. Right now, this issue around use case being use case agnostic is consistent across both cost and quality measures. And the quality measures, the criteria is exactly the same as what we have here.

So, yes, to answer your question, Cheryl, no, we don't have that. But it does seem -- I guess the question that I have here is that it does seem like the question of use seems to be present, or influences the way we think about the criteria. And that issue does seem to be consistent across both quality measures and cost measures, depending on what the use case is.

MEMBER DAMBERG: Right. But I guess, what I think applies universally, whether it is a cost measure, a quality measure, or some combination of the two is that the

measure be valid and reliable so that you are not misclassifying providers. And I think that if you have good measurement properties in place, then that should give you the latitude to use the measures in lots of different applications.

CO-CHAIR ASPLIN: John?

MEMBER RATLIFF: I think Taroon's point is very appropriate and very accurate. And not so much what are we measuring, but are we measuring what is relevant to the given frame of reference that is looking at the measurement? And what is your frame of reference in looking at these metrics, especially as we look at quality.

We brought up total hip arthroplasty a number of times. If I was a patient and I am looking at a five-star hospital for total hip arthroplasty, I would probably have a set of expectations as to what quality means. But if I look at the metrics that we follow here, it is 30-day readmission, 30-day

perioperative complication incidents, not return to function, not return to work, not return to pre-injury level of function, not the kind of things from the frame of reference that a patient would look at in defining quality.

While we were going through this discussion, I was checking my email, and I got an email from Castle Connolly, which are the guys that do American's Top Docs. And if I write them a check for $3,000, they will put me in the New York Times with a little picture saying that I am a Stanford spine surgeon. That may be what a patient approaches defining quality as.

But looking at how these metrics are going to be used, and especially as part of the Hospital Compare website, where we will be starring these systems, I think that is probably the first question that NQF needs to answer, and the first question that we should approach. And again, I think it is going to very based on frame of reference, whether it is

a payer, a provider, or a patient. I don't know that there is going to be one answer to the appropriate use definition.

CO-CHAIR ASPLIN: We'll passing the hat for John right before we leave to see if he can win that issue.

(Laughter.)

CO-CHAIR ASPLIN: Ariel.

MEMBER BAYEWITZ: I don't know if I would change how we evaluate measures. I would say just to comment, it would be helpful if, for the developers, in selecting the measures and prioritizing what they put forward, that they would take into account the use.

I think the CMS measure that was presented, clearly, they had thought through how it was going to be used. It was going to align with this other measure. It would be used to push hospitals to function in a certain way. This is how, potentially, it would be displayed. So, that was in the back of their mind and they put forth the measure. And then,

so that are other points are made, then we can evaluate is it valid, is it reliable. And that is the job of this group.

I think the other two measures with the RRUs, the measure itself was focused on plan comparisons. And again, I think there is talk about, theoretically, one can learn from it, and it is a little bit amorphous. And there is value in putting measures out there that are amorphous that we can look into the detail and figure it out and it is not forcing down on the high on us.

But I, personally, there is just so few resource use measures out there right now, it would be helpful if when the developer is pushing for the measure, they have in mind this is the specific -- this is a or a few specific ways the measure could be used. This is the specific person, or body, or group that is accountable for driving the change for that measure. And if they had that in mind, I think it would be, I personally think, we would get

more benefit out of it. And then again, we could evaluate is it valid, is it reliable, et cetera.

CO-CHAIR ASPLIN: Matthew.

MEMBER McHUGH: I would agree. I don't think -- I think it is problematic to change the bar for reliability and validity. But like you said, there was kind of a silent discussion that we were all having, but there was clear use for other measures that we have looked at. But that information wasn't part of what the materials -- but that was kind of underlying a lot of the discussion. So, I don't think that changing the criteria are important but being open about what the use is, because it is driving the developers, in some cases.

MEMBER BAYEWITZ: Like, just taking asthma as an example, if you look at asthma. If we think the provider is the one that controls the actual resources, which they do, right, so if the measure is at a plan level,

the next question is okay, well, let's say a plan is looking at this, and says I am noticing there is a difference with how my resources are playing out for asthma, versus other plans.

I want to drill into this. I want to look at how does this look from a provider perspective. Maybe I will look at a market. Is there a variance within New York versus Colorado. And then within that, is it variations across provider organizations?

If the developer was thinking about that type of drilldown, starting at the plan and then market, organization, possibly even provider, then they would look at those metrics. They would say well, how many asthma patients are there in a population. And within that, how many asthma patients are we excluding because of our exclusion rules? And is this measure statistically meaningful at a market, and then organization, and provider level? And could you, in fact, actually evaluate organizations, so within a Shared Savings or

Risk Model, would this measure actually be appropriate?

If those kind of things were part of their thinking in terms of putting together the measure, I think we would get better measures and then we could actually use them. Right? I mean, at this point I don't know if you could actually use some of these measures. They may be valid at a certain level, but is any plan going to actually then use it to incorporate into a program to actually drive change?

CO-CHAIR ASPLIN: Nancy.

MEMBER GARRETT: So, I agree with Matt. I think the straw model here, of a different standard for payment versus reporting, it is operationally so hard to do because once the measure leaves here, it is out in the world, and we don't really know what is going to happen. And a lot of times these measures are taken up by small community groups, because it is an algorithm they can just apply. And so to be able to know all the use

cases ahead of time is really tough.

I mean, I think we are in a situation where we probably need to assume that every measure that comes through might be used for payment, and maybe that is the assumption, because that is the world that we live in. And that is where healthcare is going. And so, we should kind of have that a bottom line.

And so it was an interesting discussion with CMS yesterday, where they said this is just going to be used for reporting. But then they kind of backtracked and said but I actually can't promise anything.

(Laughter.)

MEMBER GARRETT: Did you catch that? But at the same time, I feel that we can't, it is almost like putting our heads in the sand to try and consider these measures without the context in which they are being proposed. So, you have had the experience, as a committee, of deciding whether to endorse a measure that was written to statute must be

used. And so without understanding that context, I think it is important to understand the context and where it is coming from and what the reason is.

It was also interesting yesterday. They were describing how they were planning to show this in three groups: average, below average, and above average. And that was almost like an excuse to say we know the risk adjustment isn't perfect, but we are not going to be very precise about how we display it. And so that felt like well, I don't think we can really use it on criteria, but at the same time, it is helpful to understand.

So, I think what I am saying is I guess I would favor loosening up that restriction a little bit because is so hard to talk about usability without that context. I think the context is important. But I don't know that we can formally change the criteria. I think that is more of a qualitative of judgment of the committee to help just set the

frame for how we are evaluating it.

CO-CHAIR ASPLIN: I would agree with the last several comments around changing criteria. I think it would be very challenging. It is already challenging. Just compare the sophistication and number of measures we are reviewing now to the Care Coordination Committee just three or four years ago. I mean, the bar has gotten so much higher already.

And I get that we are use case agnostic, in terms that we don't have different criteria, depending on the use case. I don't know that, really, we are use case agnostic. We are always having those conversations about what is going to happen with this thing. So, I think we are probably already applying that filter, to some degree.

I do think Taroon, to your earlier comment about when NQF embarked on the cost and resource use. I do believe it would be worthwhile, and I don't think this contradicts

what I just said, and to some extent the developers and the measures we have talked about have done this. But to really make it explicit, how will this cost and resource use measure be linked to measures of quality? And have that as a separate discussion point in the use and usability standpoint.

I don't think the converse would need to be true, because the foundation was quality, right? But you know I think a converse, an explicit description of how resource use measures go back to quality would be very fair. I don't believe that that should filter into a different set of criteria on scientific acceptability. Joe?

MEMBER STEPHANSKY: Two questions. Now, aren't we beginning to overlap with the MAP, when we start talking about use cases? So, how does that relate? And then I guess the other thing is, because I was one who struggled greatly with a particular CMS measure but I thought CMS was misusing, from a hospital

standpoint, but I still voted for endorsement, because it met the criteria.

And that is what I thought was -- the NQF criteria. And I thought that was what we were assembled to do, no matter what I thought about the CMS use of the measure. And I think, as a group, I look around and I think everybody here is self-aware enough to be able to do that. So, have you heard something different, say in other committees, where people are going to far considering use cases?

MR. AMIN: So, I don't want to speak necessarily about this committee, per se, but I think, even to Brent's point around the use case is in the back of people's minds. And if that is the case, one would assume that that affects the way you would weight the criteria.

In the abstract we have observed, broadly, across other, different types of committees, that if a measure is being used for the purposes of health plan comparisons, just from a public reporting standpoint, what we are

accepting necessarily for testing for reliability of validity testing may be different than what we would see for a measure that is going to be used for, for instance, a federal payment penalty program.

And it is only to suggest that, and it is only to challenge the committee to say that if one was to consider the criteria weighting differently for those use cases, it would assume that we should at least make that transparent, and part of the conversation in a meaningful way.

So, that is one -- I submit that as an observation. It is not, by any means, representative of this committee, or broadly, of every committee but that is an observation that can be observed.

Now, with that, I think there is also the question of how this is different from the MAP. We are not suggesting that the change of this committee, even if the committee recommended very strongly that the use case

would require different criteria, there are many governing structures of the organization that would strongly disagree that characterization. We would have to go through that process.

It is only more for us to get a sense of the pulse of the committee, to understand how much of that is actually driving the decision-making and where, specifically, is it driving the decision-making? Is it -- because there is different criteria that might be involved or are certain use cases requiring a higher bar?

And just to even characterize in some meaningful way that this may be happening and we could understand it and understand what we could do operationally about that. Or that may be perfectly appropriate, but we would have to have some, at least, transparency around it.

NQF is going through a strategic conversation about how the measure endorsement process and the measure selection process

relate to one another. The measure selection process, as it relates to the measures application partnership, does make decisions about specific measures for the purposes of particular programs, which has its own set of criteria and is much more of -- you know we ask individuals to sit at these table as the endorsement, with the endorsement work as individuals, as scientists, as people who have experience with these types of measures and applications across the broad spectrum of our work. The measures application partnership is much more representative of stakeholders representing the interest of stakeholders.

And obviously, we would think about the question of use case differently if you were even, Joe, as a perfect example, if you were representing the, putting the hat on of the hospitals, you would think differently about the question of use case. I think it would be very clear. Well not very -- I mean it would be different. Let's just put it that way.

So, that is where the line is right now. But we are exploring having some very strategic conversations about whether that line should be reexamined and how it could potentially be thought of differently in a more streamlined and a more efficient way, recognizing that use case is part of the decision-making process that the committee does think about.

CO-CHAIR ASPLIN: Lisa.

CO-CHAIR LATTS: So, I agree with all the comments that have been made so far, why it probably wouldn't be a good idea. But the one I wanted to add as well is that often, pretty much in most cases we are seeing today, reporting is a bridge to payment. So first, it is you make it public, you make it transparent, or at least you start reporting it, whether or not it public and transparent. But at least you start reporting it and then that is a bridge towards payment or reimbursement mechanisms.

And so if you use one strategy, one

measure for reporting and then another for payment, it takes away -- you know the whole idea is tracking and trending and getting that idea of what is going on. And so then if you change the measure as you are moving from reporting to payment, it is against the purpose. The purpose is to be tracking and trending, to get your arms around what is going on so that then when you do start going into payment, you have got a good picture. And so if you are using two different measures, it negates the whole purpose of it.

CO-CHAIR ASPLIN: I think Nancy said it very succinctly, that in the back of our minds, I think we kind of assume anything that we pass here may ultimately be used for payment purposes. And so that is why I think we already have a pretty high set.

And if you look at just the number of measures out there, kind of the natural history of quality measures where five, six years ago, it was if you had a group of experts

around the table, you basically could get it through. That is not the case anymore.

So, we have gotten better measures. But part of the reason we got better measures is having so many out there that we are saying, wait a second, let's get more stringent about the criteria and so forth, that I worry that if we narrow the pipeline too much as we are in this new area of cost and resource use that it will be too easy to just say no, it doesn't meet our criteria and sorry. And then we won't get better measures because there won't be enough of them out there in use to inform the conversation.

And frankly, that means the programs will implement their own measures without this process. And I don't think that is as effective.

MEMBER CLARK: Yes, so on usability and use, again, I think there are different audiences for usability and use. The one that we have been talking about is really from the

payer perspective. You know, CMS perspective, insurer perspective. You know payment, that is a payer perspective. But on the other hand, the data, if you are talking about reporting down to the provider level, that is information that a provider is going to use.

So, I guess in my mind then, how will a provider use that information? And one factor related to that would be translatability. So, in order to understand how these measures are being calculated, you know, from a provider perspective, I think you would want to know, be able to translate, all of this measure structure and all this documentation that was provided to us. But that is reams and reams of information. It seems to me like you need a way to succinctly translate how this was developed and what it means to a provider, as well.

So, that is kind of a little different spin. And I was thinking, maybe, I don't know if it would be useful to have another

category of like translatability. Or even if we are talking about you haven't talked about any measures that are patient specific but when that comes into play, if we ever did evaluate any patient-specific measures, you are going to need translatability to a patient, I think, as well.

CO-CHAIR ASPLIN: Joe.

MEMBER STEPHANSKY: Okay, my own opinion is coming off of Taroon's statement about that in the MAP you have people who are representing specific stakeholders, just parts of the industry.

One of the strengths I think of this committee is that yes, I can give a hospital standpoint but I really like the idea that I can be here as an individual. And as soon as we start talking use cases, now I have people to answer to outside of this room. And this is one of the few places for somebody who deals with power and politics on a daily basis, where you can actually have kind of a Socratic

conversation, which is rather rare these days. So, my tendency would be to, in a way, leave it as it is.

CO-CHAIR ASPLIN: Larry.

MEMBER BECKER: So, I'm sorry I missed 15 minutes of the conversation, but I really liked what Mary Ann was saying. And I wondered if, as we put measures out, whether there ought to be two constructive paragraphs or so, one that says if I am a provider, here's of all you should think about this measure and use it. And if I am a patient, if it is applicable, here is why I should think about this measure and how I should use it, and in sort of the executive summary of what all that is.

And with no black boxes, sure if you want to get into it you can, but at some level, in English, should I pay attention? Shouldn't I? When should I? Why should I? And make it simple for people.

CO-CHAIR ASPLIN: Lina.

MS. WALKER: I would like to

follow-up on Mary Ann and Larry's comment. I sort of agree and disagree with that perspective. We are asking the developers to do -- already it is so expensive and difficult to develop these measures. And now we are saying it might be nice if they could do even more, providing all the individual provider and care team data, specifying the measure so that it is -- translating it so that it is usable for consumers. Now, translation, as all of you know, is really very difficult.

You know, I like the idea of having an executive summary to say this is how we think the measure could be used for consumers or this is how the measure could be used for provider groups. But to ask them to actually drill down and provide all that other information seems to be a little excessive, in my mind.

I think that there are consumer groups out there who would be able to take the information and translate it. They have the

expertise. They can translate it into material that would resonate with consumers and similarly, provided you will have the data. Right now, the developers are struggling with just claims data. Even if they get to the point where they have claims and clinical data, they might not drill down to the level that would be appropriate for a plan, for instance.

So, I think it is great this is where we think and where we hope the measure and how the measure could be used but I think it is a lot to ask them to actually present all that additional information.

MEMBER CLARK: Just a follow-up, I don't think we were necessarily saying, being prescriptive in saying we need these reports down to the provider level and all these other slices and dices of data. No, we were just saying that that is how the payers are going to be reporting the information back. That is probably how it is going to be used. And we just heard CMS say they are providing

hospital-level reports back to the hospitals. I mean, I don't think it needs to be prescriptive. It is just that is how they are using it, I think.

CO-CHAIR ASPLIN: Ariel, Janis, Nancy, and then we are going to ask for any public comments.

MEMBER BAYEWITZ: Yes, I was just going to echo that. My intent was not to get a detailed reporting. It I just they need to be thinking about how is it going to be used. And that should inform the decision to put forward a measure. Because at the end of the day, if that measure is not going to be actionable in any way, shape, or form, what is the point? There still could be points, but why prioritize that versus other metrics?

And you could do that just by looking, like I said, how many asthmatic patients are there in a population? Of that, how many are we excluding? Generally speaking, how many patients are in a given

practice, organization? How big? I mean you don't actually need to have all that data to be able to think about those kind of pieces.

CO-CHAIR ASPLIN: Janis.

DR. ORLOWSKI: I think we're being a little bit academic. What we have to realize tonight, I apologize if this sounds crude, but we are moving from wouldn't this be nice to have this measure or that measure, to being pretty serious, big business. And Value-Based Purchasing has gone from one percent, it is now slotted to be at five percent, and we are talking about serious retrospective withdrawal of payment for a business.

And so I think for us to say let's not ask this or that, the industry will demand this to be as precise as possible. And when I take a look at charts that show that depending upon what methodology you use, you could either be ranked as a great doctor or a mediocre doctor, or a great hospital system, or a mediocre hospital system. I think that we are

naive not to think that these are critical issues that are going to get down to the very detail because this is big business. It is not just academic. You know, wouldn't it be nice to measure this or that.

CO-CHAIR LATTS: Although I think Lina's point was just that we shouldn't ask the developers to do it. That is inevitably going to happen, but it shouldn't be the developer's job. Okay.

CO-CHAIR ASPLIN: Nancy.

MEMBER GARRETT: So, I just have a question, Lindsey. It seemed as though you were implying that there have been concerns raised that the validity and reliability criteria don't work as well for the cost and resource measures. Can you articulate a little bit what the issues are?

MS. TIGHE: Yes, generally I think Taroon can probably speak to it better, since I wasn't here for Phase 2, but I understand there are some issues with applying the

validity criteria as laid out in the algorithm to the measures. So, really, we just wanted to understand if, essentially, face validity is an acceptable level of demonstration of validity for resource use measures, or if that is potentially different in this case.

MR. AMIN: And I think what I have heard from the committee, I mean on this, as follow-up from Phase 2 and what I am hearing again today is let's try to be as consistent as we can with the quality measures, if nothing else. I mean resource use measures are still in the infancy in the fact of how many there are and where the field is.

And so, the face validity standard, while it may not be sufficiently rigorous for some, is an acceptable standard at this point. And as we go forward, we will have to continue to revisit that. So, that is what I am hearing, I think. So, unless somebody feels very strongly -- I know there are people who feel very strongly about that. But that is

generally the consensus of what I have heard.

CO-CHAIR ASPLIN: All right. Very good. I appreciate the dialogue. And Operator, I would ask if there are any public questions or comments at this time.

OPERATOR: At this time, if you would like to make a public comment, please * then the number 1. And there are no public comments at this time.

CO-CHAIR ASPLIN: Thank you.

And Taroon or Lindsey, if you have any other final questions or comments for the committee, that would be great.

MS. TIGHE: I don't. Taroon, do you? Or just briefly into next steps?

MR. AMIN: I just have a few reflections just because we have a minute. I would just say that this work on cost and resource use, some of you have been with us since the beginning, for the last four years. Some of you are joining us now. The amount of effort that you contributed to this effort in

the last year and a half across these three phases of work, these are no small volume of measure materials to review and the complexity is significant, particularly when we send you Andy's report between Day 1 and Day 2 for review, while interesting for me and some, maybe is challenging to get through for some others.

But there was a lot of offshoot work that was created by this, by just the evaluation of the work. On reflection, this work that Andy has taken on with Chris, the measuring the affordability effort, which Lina and others were participating in, the episode grouper evaluation criteria. Many of you contributed on multiple of these offshoot activities. And we really just sincerely appreciate that effort.

I know we are going into a period where we don't' have clear next steps in terms of funded work. Lindsey will talk through that. But obviously, we will engage you as we

digest this material, identify additional areas that need further exploration for either authors or co-sponsorship as we go out for seeking additional funding.

And so, just a sincere thank you from my perspective for all of your hard work and obviously, to Lisa and Brent for carrying us through the last two days and, obviously, for their prior work. So, I will just turn it over to Lindsey in terms of next steps.

MS. TIGHE: Yes, so we will certainly keep you busy through the fall. Slide number 43. But just as a reminder, Phase 2 of the cardiovascular draft report is posted for NQF member voting right now. So, there will be some follow-up that we will share with you on that. It is going to CSAC during their July 8th and 9th in-person meeting.

And then Phase 3, we are going to do some follow-up work on that, as we have mentioned. We will be holding those calls to bring other committee members up to speed and

engagement in the conversation. You all aren't required to attend those calls, but it would be great if you are available, if you could join, just kind of give them the sense of the conversation, what we have talked about, what we haven't. So, we will keep you posted on those as we get them scheduled. And also, as we get voting results, we will let you know.

But this draft report, we are going to pulling it together and putting it out for comment August 14th through September 12th and then bringing you all back together for a call on September 24th from 12:00 to 2:00 p.m. Eastern Time.

And so yes, we will be in touch a lot, as always. I am very appreciative that you could join us. And a huge thank you to Cheryl. Jennifer, I know you joined the phone, too, so thank you for joining us virtually.

And I guess go USA, at this point.

(Whereupon, the above-entitled matter went off the record at 11:42

a.m.)