NATIONAL QUALITY FORUM

+ + + + +

LINKING COST AND QUALITY MEASURES

+ + + + +

EXPERT PANEL IN-PERSON MEETING

+ + + + +

FRIDAY
MAY 2, 2014

+ + + + +

The Care Coordination Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street NW, Washington, D.C., at 9:00 a.m., Joyce DuBow and Carole Redding Flamm, Co-Chairs, presiding.

PRESENT:

PETER ALMENOFF, MD, FCCP, Veterans Health Administration
STEVEN ASCH, MD, MPH, VA Palo Alto Health Care System and Stanford University School of Medicine
LAWRENCE BECKER, Xerox Corporation
DAVID COHEN, MD, MSc, Saint Luke's Health System
MARY CRAMER, MBA, CPHQ, Partners HealthCare System, Inc., Massachusetts General Hospital
JOYCE DuBOW, AARP
CHRISTINE GOESCHEL, ScD, MPA, MPS, RN, FAAN, MedStar Health
DONALD LIKOSKY, PhD, University of Michigan
TIMOTHY LOWE, PhD, MSW, Premier Healthcare Solutions

CATHERINE MacLEAN, MD, PhD, WellPoint, Inc.
JACK NEEDLEMAN, PhD, University of California, Los Angeles
STEVEN PANTILAT, MD, FAAHPM, University of California at San Francisco
KIMBERLY RASK, MD, PhD, Alliant Health Solutions
CAROLE REDDING FLAMM, MD, MPH, Blue Cross Blue Shield Association
IYAH ROMM, Health Policy Commission
MATTHEW ROUSCULP, PhD, MPH, GlaxoSmithKline
ANDREW RYAN, PhD, MA, Weill Cornell Medical College
DENNIS SCANLON, PhD, The Pennsylvania State University
JEREMIAH SCHUUR, MD, MHS, FACEP, Partners HealthCare System, Inc., Brigham and Women's Hospital
JEFFREY SILBER, MD, PhD, The Children's Hospital of Philadelphia
ALAN SPEIR, MD, Society of Thoracic Surgeons
JOSEPH STEPHANSKY, Michigan Health & Hospital Association
CHRISTOPHER TOMPKINS, PhD, Brandeis University
HERBERT WONG, PhD, Agency for Healthcare Research and Quality
GREGORY WOZNIAK, PhD, American Medical Association
GARY YOUNG, JD, PhD, Northeastern University

NQF STAFF:

TAROON AMIN, MA, MPH, Senior Director
HELEN BURSTIN, MD, MPH, Senior Vice President for Performance Measures
VY LUONG, Project Analyst
ERIN O'ROURKE, Project Manager, Strategic Partnerships
ROBERT SAUNDERS, Senior Director, Strategic Partnerships
ASHLIE WILBON, RN, MPH, Managing Director, Performance Measurement

ALSO PRESENT:

AMER HAIDER, Doctella.com
REBECCA HANCOCK, American Academy of Ophthalmology
DIEDTRA HENDERSON, The Institute of Medicine

A-G-E-N-D-A

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

9:06 a.m.

CO-CHAIR FLAMM: All right. Good morning, everybody. If we could all come back to our seats we'd like to go ahead and get started today.

I want to welcome everybody back to day 2. And reflecting on yesterday thank everybody for the tremendous amount of work, good ideas and energy that we brought to the table.

I heard several of you also sharing that we were all a little exhausted last night. So hopefully we had a good night of sleep and we're ready to tackle the challenges today.

I want to also thank the NQF team who's done a great job yesterday and last night and this morning in pulling together a recap. It's forming as we speak and I think it's going to be very helpful.

And I also thank Chris and Andy for

all of their leadership and work yesterday.

So, without further ado what we're going to do this morning first is recap and provide a little high-level summary which is in and of itself a little bit of a challenge of the great discussion that we had yesterday.

And hope to -- since we made up some time and did all of our breakout session presentations yesterday we're going to have a little bit of extra working time this morning to make sure that we're getting clarity and some details in places where we might them as much as possible even though there's a lot of big picture and new thinking that we're doing.

And this is such new work that we're doing that we're figuring it out as we go along. And I think that's just an iterative -- an understandable process that we're in.

Ashlie?

MR. AMIN: Actually, Ashlie stepped

out for a minute so I'll take the floor. Realtime adjustments.

So yes, again, thank you all and welcome to day 2 of this linking cost and quality session.

So, the agenda for today is that we will be going through, as Carole pointed out we'll be going through just a review of the goals and kind of the high-level summary at the risk of potentially being a little bit overly simplistic in terms of all the conversations that we had yesterday. But just summarize across the different workgroups some of the information that we've found.

And then have some discussions about general points that we identified yesterday. And then just go back to Chris and Andy to see if there are any specific areas that we might want to have some additional conversation this morning.

We'll have a break at 10 o'clock and

then we'll go at 10:15 into a discussion around implications for the NQF endorsement process. And then some summaries of recommendations and identifying the path forward.

So, just to restate the goals of why we're here. What we really wanted to do over the two days is explore the current approaches and applications for linking cost and quality both at the measure level and at the programmatic level.

And Chris and Andy's environmental scan helped set the foundation for that work in addition to our conversation yesterday around the potential challenges of thinking -- that more at the programmatic level and at the measure level we had a discussion yesterday around the challenges of thinking about linking cost and quality in a single composite measure.

Secondly, we wanted to identify key methodological challenges to cost and

resource use measurement. And thirdly, defining key principles and best practices for linking cost and quality.

And the way that we did that yesterday was to think about through the lens of four different use cases and think about considerations and potential challenges for linking cost and quality in four different use cases.

And the fourth which is what we'll spend most of today thinking about is thinking about providing operational guidance to NQF related to the endorsement process and the selection process and recommendations for application for the evaluation of efficiency measures going forward.

And I know that we had a number of conversations yesterday about potential challenges of even getting down to the measure-level programmatic-level guidance, but we want to be able to explore that in a

1 little bit more detail today and understand
2 what really the path forward is for the
3 endorsement process and the selection
4 process.
5 And as a few of you noted yesterday
6 while there are some conceptual challenges
7 to being able to do this there is the
8 practical reality that NQF is being asked to
9 do this work currently and this in some ways
10 is already moving.
11 So we want to be able to make sure
12 that we're doing this in the most
13 scientifically sound fashion and that we're
14 advancing the goal of improving healthcare
15 value ultimately but at the same time not
16 sacrificing healthcare quality in the
17 pursuit of monitoring healthcare costs.
18 So ultimately that's the goal of
19 what we're trying to achieve today. We made
20 significant progress yesterday so I thank
21 you all for a pretty exhaustive day. It was
22 very good, deep conversations. And I hope

we can continue that progress today.

So, does anyone have any questions about what we're trying to achieve today or any general reflections about the discussion or the conversation over yesterday? Okay.

So, we're going to go into just a brief discussion of some general points that we identified yesterday. I'd welcome, Erin, for just a quick high-level summary of some of the main topics and themes that arose yesterday.

And then I'd welcome some comments from Chris and Andy about any general reflections of the high-level themes that we discussed yesterday. And then would open that up to the committee. And then we'll go into more of a deep discussion around what the breakout groups identified. And so, Erin.

MS. O'ROURKE: Thanks, Taroon. So, a few of the high-level themes that we really heard yesterday was first that we

need to clarify the relationship between quality, efficiency and value. And before we can really move forward it's necessary to know what exactly we mean by each of these terms.

We discussed that composite measures may be mathematically possible but are probably not feasible or practical for real-world applications.

We decided the selection of a model may depend on the audience and the goals of a program. Consumers and payers want information to help them make informed decisions and to really know where to take their business if you will to get the most bang for their buck.

But providers need granular actionable data. A summary score that might be useful to a consumer won't provide an organization with the data they need to see where they have inefficiencies or where they can improve cost or quality to become more

efficient.

We outlined a few challenges to implementing efficiency measurement. First, a question about what level of scientific rigor is necessary. Different audiences may need a different level of detail.

There's also a question that we started to raise yesterday about whether the models should be reliable and valid within themselves, or if it's necessary to show that you would get a similar ranking if you applied different models.

Finally, we started to discuss that data may be a challenge to efficiency measurement and the current limits to cost accounting may impede cost measurement. And may limit providers in what data they have available to really drive down and improve their costs.

MR. AMIN: Erin, Vy, do you mind putting up the conceptual model for a second?

There's one thing that I just wanted to walk through again. We had a bit of this discussion during the web meeting a few months ago. But I wanted to just remind the group.

I know we had a lot of conversations yesterday about definitions and how we're thinking about this group. And the question of sort of objectivity versus subjectivity and where that sort of comes up.

So the way that NQF thinks about this work is based on the patient-centered episode of care framework which was really the foundational piece of work that identified that we need to really move toward a measurement framework that's looking at cost and resource use in conjunction with quality across time.

And ultimately there are a number of different -- and where some of the definitions first came from is really that original framework.

And so the work around cost and resource use is really -- the way that NQF thinks about this is that they're really counts of utilization that are generally monetized but don't necessarily need to be. So, counts of utilization monetized either using actual prices paid or a standardized pricing approach.

The unique feature of how we think about cost and resource use is generally there's not a clear indication of directionality. We don't necessarily know that up or down is actually better or worse.

And so to really understand health system efficiency we really need to pair the information about cost with quality to be able to understand some sense ideally -- and the way we've been thinking about it, and I think that's open for some discussion, is that there's an objective evaluation of quality -- or efficiency that pairs quality along with resource use.

And so this is sort of the orphan that Chris described yesterday in some ways, that there's this middle ground between value which really thinks about stakeholder preferences of the weighting of cost and quality and sort of the pure resource use measurement.

And so this is the area, this sort of second-ring box here around measuring efficiency is really what we're here to discuss, whether there's some objective way to look at cost and quality together without the question of really getting toward value which introduces the preference element.

DR. SCHUUR: Why is time there? I understand early on there was a mention of sort of time to see providers as a measure that people have labeled as efficiency but that's a timeliness measure. It's sort of inappropriately labeled an efficiency measure.

MS. WILBON: Yes. I mean, actually

I would probably -- I would say let's ignore that box for now.

I think in our earlier stages this model was kind of looking at time in terms of some more of an activity-based costing type thing. But I think I would recommend for this discussion let's just kind of take that off of the -- pretend we don't see that box right now.

But for the purposes of this discussion kind of looking at the cost and resource use with the quality in terms of thinking about efficiency.

DR. SILBER: A question about the definition of the cost again. Because it depends on the stakeholder, right. So, cost for the individual patient, Lawrence was saying true north, is really what they pay. I don't think they care too much about whether the hospital did it in the most efficient way. They want to know what their premiums are for their insurance policy when

they join whatever organization they're going to join for their care.

So, then there's cost for the hospital which is they're really interested in what resources they need to buy to provide the services to get the quality that they deliver. So there is bean-counting.

Then there's payments to Medicare. They're a little bit less interested in bean-counting. They're more interested in what they -- whether the payments that they give out are providing the quality that they would expect.

So I think we should at least in part be clear to define what we mean by costs in terms of resource utilization, payments by patients, payments to providers from Medicare. Because all of those are going to be very different depending on the stakeholder.

MR. AMIN: And I would just clarify that the conceptual framework from which

this is based -- knowledge is that there's various different sort of costs that are borne by different parts of the healthcare system.

And depending on the level of analysis as you're describing, Jeff, you would count different types of resources to make sure that they're reliable and valid.

It's only to sort of orient us in terms of the nature of where these terms arise. I know many people think about efficiency in a different way. This is sort of the precursor work from which this committee has been convened.

So I want to make sure there was a sense of perspective. It doesn't mean that you necessarily have to agree with. I just wanted to provide that sense of perspective of the panels that have come before us.

MR. BECKER: So, I know that you're saying let's ignore time on that chart. So let's imagine that it's not there for a

moment.

I would not break apart efficiency and resource use the way you've got it. I mean, resource use and quality are part of efficiency. So, when you display it that way it's almost as if those resource use and efficiency are two entirely different things.

MS. WILBON: It's more that the box subsumes both of them. So if you think of it like an inlaid diagram.

MR. BECKER: I think I know what you're trying to get at.

MS. WILBON: Okay.

MR. BECKER: But the way this is laid out it's sending a different message.

MR. AMIN: All right. So I just wanted to provide that as sort of background in terms of how these concepts relate to one another in terms of the background for this work. But I just wanted to maybe transition from this conversation which is just to

provide a little bit of background.

Before we get to the summarization of some of the breakout groups I just want to turn it over to Chris and Andy if you had any specific sort of overarching comments from yesterday's discussion that might set the tone for the next hour and 15 minutes. Then we can go into some summary of breakout groups from yesterday.

DR. TOMPKINS: Well, I hope I don't necessarily set the tone.

(Laughter)

DR. TOMPKINS: What I'd like to do is consolidate some gains, maybe, or at least put out some refutable assertions about what I think might be emerging.

Because one of our collective -- as much as we might agree and disagree, and I think we mostly agree, we're trying to come up with a consensus statement among the committee members and authors and NQF staff so that you can move forward with something

that you have some confidence in that you can say the committee actually stands behind whatever this paper says and whatever NQF turns into public statements about it.

So, just a chance while it's still the beginning of the second day we can revisit this and change it so that Andy and, for example, don't go off the next several weeks and write something that really doesn't do that. So I welcome actually participation in this particular process to do this right now in this moment but also for the rest of the day.

Yesterday for me was fun and interesting, especially fun because I wasn't necessarily bottled into one use case group. Although people who were bottled in probably had the most fun of all because it became deep and engaging. And I bounced around not to all of them but to some of them.

And for me it sort of represented a microcosm of the world because in this room

yesterday was a reenactment of a microcosm of the world. We have people in one corner who are trying to focus on QI and they're trying to grasp for target opportunities and actionable information.

And they try to say we have to understand that going forward we have a job to do and we have to consider both of these things simultaneously.

And then we have this group over here scheming about network design. How are we going to get rid of the bad apples. Where are we going to draw hurdles. Who's not going to get over it and so on.

And the p-for-p, how are we going to get these rascals to change their mind. And then public reporting, how are we going to get consumers to stand up and pay attention and use this information.

It's sort of an interesting separation of perspectives which of course was by design, but it was also fun and

interesting to observe as it was happening.

One of the things that I think cuts across all of that. So here's sort of like one statement that I think the group would be ready to jump onto, but again, asserting it to make sure that's the case is that everybody needs to acknowledge that they're working in this space.

Everybody needs to be here. The day of quality is over here doing its thing, quality measurement is over here doing its thing and over here resource use is doing its thing, it's -- we're in an era where the quality improvement has to be cognizant of resource requirements. And making judgments about resource use and so forth has to be cognizant of the different outputs that are. So, everybody has to be in here.

Now, in here means just sort of operationally for a minute one of these models, the models that we've enjoyed talking about so much for the last day.

None of them really fell out to say that, you know, this model -- to play in this space this model simply doesn't have a role. There's a role for each of the models. So therefore we can put forward to say that to actually work and play and be serious in this space can involve participating in one of these models.

Now, you should do it with your eyes open. You should understand you're choosing a model that's different from the other models and there are tradeoffs and so on. But all the models in fact are legitimate for some purposes at least.

One of the observations that I made sort of yesterday but also during the environmental scan is that there can be hand-waving.

That is to say, I mean even written in the environmental scan writeup there's, you know, we've got this grid and the columns and so forth.

And then you come to one row and it says well, what is their measurement of resource use. And it says, quote, "are they working on their cost," something to that effect. And it was like, well, they sort of get it, they're working on the cost and therefore they are. What does that mean? Therefore they're efficient? Therefore they're using resources wisely?

And then in the conversation yesterday there was a similar allusion on the other side that you might have a program that's set up that's really around payment reform.

And then along comes the fig leaf of the quality measure. You know, well, we have quality measured as a structural measure and the structural measure is does the hospital have a roof and four walls. You know, they say okay, yes, quality met. Game on, let's talk about resource use at this point.

So I think that in addition to saying that these models are perhaps legitimate in their own right there has to be some specification about what does minimum participation mean.

And we can't just say -- I think in other words we have to call out this notion of hand-waving and to say something about -- even if it's side by side which for the moment might be the simplest thinking model because it doesn't have the concerted attempt to put them together.

Even there to say that, yes, we're participating because we're doing side by side, there has to be some rigor, some minimum floor in terms of what participation really means.

Now, one of the things, again, sort of a generalization from yesterday is that everybody is involved in these decisions. Sometimes they're implicit and sometimes they're explicit. That's one dimension.

Another dimension is that sometimes decisions are made just for -- by individuals for themselves in which case we think, I think, again, refutable, that in circumstances where the information is generated for the purpose of individuals to make decisions for themselves we can allow maximum discretion.

That is, for example, side by side is maybe fine, maybe good, maybe minimum for things like individuals who are trying to make choices. Because this question about how do you weight this and some of the factors aren't there like travel distance and so forth. We don't want to corner people in with presumptions or baking in assumptions in terms of what's better and what's worse. And so therefore when the individuals are really left to -- when decisions about that, then maximal -- emphasize simplicity, transparency, accessibility and then the flexibility of

being able to make decisions like that.

But, and it was interesting going around the group how there was acknowledgment, sometimes said, sometimes implicit, that we're actually making decisions that will affect lots of people all at once.

And in one of the group I was sitting in it rolled all the way up to societal level. Hence the reference to death panels and so forth. That there is actually explicit tradeoffs going on here that can affect whole classes of people. Whether it's you're too old to have that operation, or our budget can't afford that at this time. So there's a lot of actually hard tradeoffs that are being made.

And sometimes, well society up to this point most of all have just sort of floated along and let that be implicit. But if the science is going to start marching forward and thinking about how is it not

only for individuals making up their own minds, and not even for health plans making up some decisions that are systematic and affect a lot of providers in their network decisions, but even rolling up to higher levels where everybody is affected at once.

In other words, the more that there is a systematic effect where a decision about this necessarily affects whole sets of providers in terms of who's in, who's out, who's in the network, who's not, or who wins and who loses. And eventually who can get care and who can't get care.

Those decisions actually should be made in a way that is more rigorous, more objective in the sense that the factors that go into the decision-making process are actually asserted and can be debated, and the methods that use those objective factors should in fact be rigorous.

So, the models -- even though the first sort of plenary statement I made about

all the models are legitimate in their own right, to use them not only requires minimum but to the extent that you're involved in the decision-making processes that are more systematic you actually should move beyond the minimum and move to models that are more discoverable and replicable and eventually objective, and methods that treat with sort of rigor and respect the necessity of including in those decisions that come out of those models as much of the explicit tradeoffs as there are involved.

DR. SILBER: Can you define legitimate?

DR. TOMPKINS: When did I use legitimate? I'm thinking extemporaneously here. What was the sentence?

DR. SILBER: You were saying models are legitimate. You used the word "legitimate" a number of times and the models being -- and I don't understand.

Because the NQF, up to now if you're

-- generally been certifying measures relating to quality there isn't this kind of debate. It's when you put these together that this becomes very different.

So I want to understand what "legitimate" means. It might be that the NQF says it's not legitimate to form a function that combines quality and cost. So I just want to understand what "legitimate" means.

DR. TOMPKINS: I guess -- I didn't intentionally use that word. I guess what I meant was that if we're going to make a statement -- it was really drawn off my first statement which is that NQF might be in a position to put a stake in the ground that says it is a requirement of the measurement science world to be cognizant and even act as if they were cognizant the fact that the use of measures should be multidimensional including the cost and the resource uses at the same time.

And that maybe my first use of the word "legitimate" was to say that the discussion yesterday did not lead me to think that any of the models that were discovered in the environmental scan either in play or in the literature were not at least an acknowledged attempt to do that and therefore are legitimate.

CO-CHAIR DUBOW: If I think about that in the context of what NQF does to make things legitimate there are explicit criteria for scientific evidence, for feasibility, a whole host of very specific criteria that have been accepted by a consensus body.

It seems to me that Chris is suggesting that that kind of legitimacy because it suggests acceptance by a consensus body is the direction. If I understand you correctly, Chris, that's how I would interpret that with respect to its applicability to NQF. That's what makes it

legitimate. It's broadly accepted. But the reason it's accepted is because we have very, very clear criteria to assess and to make judgments about.

DR. TOMPKINS: I have a question actually to Taroon in your introductory remarks.

The title of this entire exercise is "Linking Quality and Cost." And one of the results of that can be characterized as an efficiency measure. And then another result of that can be a value measure.

Were you meaning to say that the committee is saying that all of this connecting or linking of quality and cost is specifically and only for the purpose of defining efficiency measures? Okay. And not value measures?

Because I think clarifying the difference is actually part of the role of this.

MR. AMIN: It is and I --

absolutely. So I don't want to just give you one-word answers but I would say yes. And it is part of the scope.

Maybe, again, a sense of perspective here would be helpful for us and for the group.

You know, the main question and the struggle that we have at NQF is that we started down this journey on looking at evaluating and measuring cost measures.

And we know that these measures are being used potentially appropriately or inappropriately in conjunction with quality measures for programs.

Now, the question is that we don't -- the science of how that is used in programs is still new. And the most important thing that we don't -- that we want to make sure is not happening is that you're just purely comparing on cost with the underlying assumption that quality is equal in the system. And we know that

that's not true.

So that's the history and the foundation of what this organization is intending to do. And so for the purposes of ensuring that we're looking at cost which is an important imperative in its own right we don't want to undermine the quality enterprise.

So ultimately putting these things together, linking them if you will, can -- if this idea of efficiency as an objective concept is not accepted, or it needs to be explored I guess. Let's leave it at that.

DR. TOMPKINS: Right, because that's one of the things that I think is purported to be the difference between the efficiency and value.

But I would say two statements. One is that starting with this basic concept of linking cost and quality measures the only real reason for doing that is to go to value.

And then the question is what's the utility in the meantime of the way station of efficiency. And where efficiency has this specific definition that calls for the evaluation of cost in the particular context of the specified level of quality.

It's like what I said yesterday that you're making determinations, quote unquote, about efficiency first with regard to understanding the specific level of quality you're talking about. You just don't mix all providers together regardless of the quality level, come up with some common cost benchmark that applies to everybody and then saying, well, yes, you may be high-quality, that may be justifying high cost, but we don't look at it that way. We just think that you're high-cost.

The efficiency measure is to say before you jump to the preference weighting do you first want to stop and make these objective evaluations of relative resource

use in the context of specific or specified levels of quality.

But even if you do that it's so that when you combine the resource use and the quality part of it what you're doing is you have the more sophisticated measure of efficiency which reflects the resource use when used in combination with the quality measures it becomes the preference weighted value measure.

CO-CHAIR FLAMM: So I think this is a very helpful and important clarifying point.

We have three committee members who would like to make some comments. I'll go Cathy, then Jack, then Joseph.

DR. MACLEAN: Chris, your comments just now and I was kind of reflecting also over the evening last night. I think we maybe want to explicitly call out where this body of work ends and where comparative effectiveness begins and the relationship.

Because, you know, the death panel discussion which we were having which was really if you were to think about the outcome in terms of, say, a quality-adjusted life year.

And the discussion came up around cardiothoracic surgery and should you do a -- what's the value of doing a valve in a 95-year-old versus a 60-year-old. And that's starting to get really into value judgments and into comparative effectiveness.

And I just -- that's fine but I just think we want to -- I think when most people are thinking about the efficiency component they're not thinking in those terms. I think they're thinking about the same outcome, you know, the outcome's equivalent and then how does the cost differ for that outcome.

DR. TOMPKINS: Yes, I mean it's a little bit -- I don't want to subject people again to this image but it's what I was

referring to yesterday when I referred to the pinwheel. Which is that this body of work is really concentrated on some set of more basic and definitive issues.

But it actually -- you can see how you can draw lines from it. To say that this concept that maybe network designers are grappling with or public reporting actually has a societal level part to it.

And it's been usually the case that the 95-year-old woman would get the valve given the right insurance. And/or if you're in a health system that has a big endowment and sort of isn't pinching pennies and so forth.

But then you put the same person in a different context maybe without the insurance, or maybe in a hospital that doesn't have resources that they can subsidize the treatment for and they don't get it. The question is is that okay.

But that's one of the spinoff

things, to say that to the extent that this linking quality and cost even raises to the point of societal benefits versus resource allocation to healthcare is -- what we're talking about in the body of this extrapolates to that level. And so therefore we might want to think of an orientation that ultimately rolls up to that level as well. But without getting into the death panels.

CO-CHAIR FLAMM: Go ahead, Andy.

DR. RYAN: I think that comparative effectiveness idea isn't -- it's thought to be generally applied to a population of patients that this treatment, you know, is effective, it's cost-effective, it's more effective than something else. Typically the statements aren't made with respect to an individual provider. It's kind of like this treatment is working on a population level.

But you could think if -- in the

context of this work you could imagine that certain procedures would be cost-effective if done by certain providers that had a certain cost and quality profile.

So I think that is a very kind of interesting extension of this, that it would -- conceptually at least you could get to a more kind of fine-grained concept of comparative effectiveness that wasn't just at the kind of population level but was brought to the provider level.

Now, we're not there at all, but that's how I've kind of thought about that kind of question. There's some things that would never be comparative effective, some things that would be cost-effective regardless of who did it, and then some things in the middle that, you know, it would be cost-effective if some set of providers did that procedure.

So, you know, I think mapping these ideas on -- connecting those concepts is an

interesting extension of this work. But the differences I see are those that I just mentioned.

DR. MACLEAN: Cost-effectiveness could also vary depending on who got it, right?

CO-CHAIR FLAMM: So, Jack?

DR. NEEDLEMAN: Okay. At some point this morning I would like us to have a conversation about whether we really have seven models, or three, or three and a half models here. Because I think it will affect the way we think about this stuff. But I don't want to do that right now because that will take us off in a completely different direction.

I think this issue of efficiency relates back to one of the summary points that was made that Jeff has pushed quite appropriately and very heavily about whether we want a single measure.

We always go back to the car

analogy. Don't ask me why we always go back to the car analogies but we always go back to the car analogies.

You can be efficient in building a Yugo. Right? But that's not the car we want to drive.

So, we can have highly efficient production of low-quality services. Lowest possible cost to produce that service, absolutely. And that an economist would say is efficiency.

So, throughout our conversations there's been this concept that there's this minimum level of quality that we're prepared to acknowledge, we're prepared to accept, and there may be a quality level that we aspire to, and that we want the system pushing itself to.

And the measure should allow us to measure where we are and also enable us to understand where we need to go.

So if we look at -- so this was one

of the reasons why I think our committee and some of the others have said we don't want a single measure of efficiency because it doesn't capture where you are on the quality spectrum.

And we want to know where you are on the quality spectrum and how much it costs you to get there. And are you efficient in getting there.

And when we look at the data from the measures that have been presented from the data that's in some of the papers that we've seen there are two cases here and each one needs to be analyzed separately.

One of them is that case where we see high-quality low-cost providers. That space is filled. There are a number of examples in there. We know that's aspirational. We can both improve quality and get to lower cost.

And in those cases many of us would argue we want payment systems that encourage

us to move everybody into that space. Raise your quality if it's lower, lower your cost if it's higher.

There are other cases when you do that mapping of cost and quality we don't see a lot of -- we don't see anybody, we don't see a lot of folks in that space where it's low-cost and high-quality.

There is this tradeoff. To get more you need to spend more. And there the issue for individual patients, for payers acting as their agents is how much more quality is worth spending that much more money on.

That's a different question than, gee, we see examples of low-cost high-quality places. Everybody looked like these people.

So we've got these two very different situations depending upon whether we think low-cost high-quality care is feasible. And the measures should allow us to determine what that space looks like to

figure out what we're trying to do here.

The comparative effectiveness stuff if we've got the clear tradeoffs between cost and quality or cost and outcomes. And the aspirational if we've got the case where low-cost and high-quality are both feasible.

So we need measures that enable us to understand what the space looks like, to understand what's feasible, to understand what decisions we have to make and what we should be aspiring to.

And the question is do these methods enable us to do that and talk about it in a way that is thoughtful, intelligent, supports the decisions of payers, supports the decisions of patients -- that's where I should have started. Supports the decisions of patients, supports the decisions of providers acting as the patients' agents, supports the decisions of payers acting as the patients' agents.

But efficiency by itself we don't

want to have efficiently-made Yugos in healthcare. We want something better than that. Even if it's efficient.

And I think that's the reason why the argument is for at least two different measures are keeping the quality measure always there even if you're measuring how efficiently the care is being delivered.

CO-CHAIR FLAMM: Joseph. Then I'll go to Iyah.

MR. ROMM: So I find myself going after Jack again which means I have very little to say.

I guess the only comment that I would make is, Taroon, in your frame I think that part of what I keep getting stuck on is you frame sort of the role of the NQF in this as being in the space of we know there's variability in quality and people are looking only at cost.

And I guess I would suggest even more so than -- and to Jack's point, even

more so than thinking about measure frameworks that help us solve that problem there are very different structures.

So in Massachusetts as we have increasingly folks coming onto risk, we have risk-bearing provider organizations that now have to go through solvency assessments. And we have mechanisms for patients who are in risk, in various risk plans that have an ability to complain about lack of services.

That to me is a very, very different construct than what we're talking about here. And really the only useful tool of all of these conversations in the space of assessing whether you are receiving care of sufficient quality services is side-by-side measurement. And I think we can move on if that's the exercise.

I think as Jack has framed there is another very interesting and exciting space for this group to start to delve into in thinking about for a variety of other

purposes where this measurement frontier goes.

So I guess I would encourage that if that is sort of the range of the construct we can answer that quickly. But I think we're trying to answer all questions with one solution at this point and it's getting us into trouble.

CO-CHAIR FLAMM: Peter.

DR. ALMENOFF: I think for the white paper the area that I really haven't heard -- I keep hearing the word "efficiency" and "quality."

And quite honestly when you talk with providers those don't go together. Efficiency is purely about money and efficiency. Quality has never looked into that. In this group we seem to be putting these words together and that's not what the public perceives as efficiency.

So if we're going to talk about quality and efficiency, or if we're going to

talk about efficiency with quality in it I think it's a whole new re-definition and education process because the average provider, the average layperson will not put that together. So I think that's one area.

The other thing is that efficiency generally is a really dirty word in the world. And I notice that when you try to get a message across or try to push a metric it's all about the message.

So, you might be saying -- you might be talking about the same thing but you say it a little differently. You know, when we had a roundtable a couple of years ago with Medicare we decided that the good word was waste reduction, not efficiency.

Because nobody can argue about waste, but everyone can argue about efficiency. So I think waste should at least be described somewhere in this paper. Because it's a really easy -- you know, no provider can push back and say we want more

waste. It's really hard to say that.

Whereas it's much easier to say well, efficiency, you know, the Yugo concept of we're going to have the Yugo healthcare system and nobody wants that. So, but nobody really can argue about the concept of waste.

So just, the two points of if we're talking about efficiency. In my mind when I got here I don't look at that and quality together. So to me they're two separate items. Maybe everyone perceives that differently than I do.

And then the second piece is we really need to at least address the word "waste" because I think that's a good term that the public and the providers understand pretty well. And they really have a hard time pushing back on that.

CO-CHAIR FLAMM: Thank you. Okay, Dennis.

DR. SCANLON: So, following up on

Peter's comment. I agree that waste is something that maybe deserves a little bit more attention.

But in terms of the public or consumers understanding or not understanding efficiency I'm not sure that I would agree with that. I think that -- and my view would be that efficiency is sort of holding quality constant.

If we think about sort of other areas of our lives, school district funding, for example, I think efficiency is a concept that consumers understand quite well. I want a certain quality education in my community, whatever that might be, but I want it efficiently. I don't want the school district wasting money.

In higher education this is all the issue right now is tuition and a high degree of tuition. How can we gain more efficiency but still maintain that level of quality. So I think a kind of understanding of

production and pricing that is competitive in a fair deal relative to a level of quality I think is something that consumers can grab onto.

But I do think it points to sort of this issue of, again, what is the purpose and how do we expect realistically people might, or organizations or entities might use this information.

When it comes to consumers the only rationale I can think as to why consumers ought to care about efficiency under traditional insurance arrangements is because of what it does to kind of the overall sort of cost of insurance generally. And that's something that doesn't really enter in any one individual decision.

When I'm choosing to go to this provider or that provider because I'm only paying a fraction of the care. So what matters at that point in time to me is price. And price is not cost nor is it

efficiency.

I can have a $20 copay for very efficiently priced care or very inefficiently priced care, or produced care. So at the point of service, and there's a lot of move to kind of give consumers price information.

But I think it's important to make sure we understand that price, or out-of-pocket price is not the same thing as cost which is not the same thing as rendering a judgment regarding efficiency.

And so for consumers it would seem to me that packaging information in a way that says, hey, just like your school district is not an efficient producer of education, your health plan is not an efficient producer of education.

In your community there is variation in terms of sort of those who are if you want to use waste, wasting resources, or those who are sort of using -- producing

healthcare more efficiently.

That becomes a different message and a different level of presentation than I think what a lot of people sort of push for which is side-by-side information for each provider.

CO-CHAIR FLAMM: Okay, Jeff and then we'll go to Taroon.

DR. SILBER: Maybe -- I feel a little bad because we're kind of saying we don't want to do what Chris and Ryan want us to do.

But actually I think what the feeling of the group is is that you have to break these apart before or during the presentation of doing what you want to do which is either say what's quality conditional on cost, or what's cost conditional on quality.

I have a sense that those are the three different ways to do this, that maybe what the consensus of the group is, and I

could be wrong, is that if we're going to do this thing which you say is legitimate which is give me cost conditional on quality, or give me quality conditional on cost in some fancy model like you referenced, the Timbie model and others, right?

If we're going to do that what we're saying is you have to also give us the side-by-side. I think that might be what the committee is feeling is that if you're going to give legitimacy to the conditional models you have to break it apart. Maybe that's just one point.

And then in terms of the details of the conditional models which -- the elegant model that you reference a lot, Timbie and Normand model which is used for a couple of your different cases.

We haven't -- I don't know if we'll get to it today but we haven't got into the details of whether those models should or shouldn't be used. Then you get into all

the issues that Jack had talked about before in terms of shrinkage, et cetera.

So, maybe we haven't achieved everything that you want from us in this meeting but I think in part it's because we're pushing back on the combination of cost and quality.

But maybe what we're keeping up with is this idea of if we're going to go that route there are some prerequisites that are needed in any presentation of the case.

CO-CHAIR FLAMM: We had wanted to have this session end at 10 and you need to present a slide, Taroon. Do you want to let --

MR. AMIN: I would say keep going.

CO-CHAIR FLAMM: Okay, all right. Andy, did you want to make a comment in response to that? And then I'll come over here.

DR. RYAN: Yes, I did. And you know, Jeff, his comments were as if we'd be

disappointed by that comment. And I've had numerous people express condolences for yesterday's session.

(Laughter)

DR. RYAN: But you know, I thought it was really -- this has been a great conversation, it's been very enlightening for us.

And I agree with what Jeff just said and I think something that Jack said. And I'll try to summarize it and maybe they can disagree with me.

But when we're talking about this linking cost and quality it seems like there's no reason to not do side-by-side. So, if we just take that, okay, let's start with side-by-side.

Let's come up with an intelligent way to combine the two domains not to develop an efficiency measure per se but an efficiency profile which is kind of the output of these models we're talking about.

And then let's subject these profiles to the kind of testing we're used to at NQF around reliability and validity to say -- to see which one holds up under which circumstances, which one seems to make sense.

And so given that it seems like there's almost -- I feel like once you get there there's almost no reason to develop an efficiency measure. Because you're kind of losing information.

If you keep quality and cost separate you have a method to combine them for a profile. And then we categorize providers given the output of the model. I don't know, it seems like that's -- there's some -- I'm going to assert that there's some consensus around that.

Now, there were a lot of other general issues that were raised yesterday about public reporting and how public reporting can best be used to maybe engage

patients that I think are -- I think there's particular use cases that we need to think about the efficiency measures.

But some of the issues I think raised yesterday, they're worth invoking in the paper but don't -- and for commentary's sake but we don't really need to -- we can't really make a decision about those.

Price I would say, price standardization kind of depends on the circumstances as well. We should comment on that, provide some context, but I'm not sure about what we can do in terms of pushing recommendations.

One thing that I feel like we didn't get to yesterday that would be useful to hear from the group and also maybe NQF is I think there's kind of an idea -- so, and we'll talk about endorsement later.

But, you know, the measures, the input here, the inputs, I think there's an idea that they should be NQF-endorsed on the

quality side or NQF-endorsed on the cost side.

But what we've seen in the scan, the composite measures are just people come up with different ways to create composite measures. And I don't think I've seen any kind of NQF-endorsed composite measures on the quality side that have gone into efficiency.

So I don't know if we want to make a statement about how the composite measures should be constructed on the quality side, if that's something that -- or we should leave it up to the developer, or we should have some guidance there.

Because I think that's an important piece of this, of that kind of sausage-making process is how you get to the kind of quality metric which is crucial. And I don't want this to kind of degenerate into all the problems with composites or all those issues, but I just wonder if we should

just remain agnostic about how that's done or you know, have some guidance, suggest some way forward with composites.

CO-CHAIR FLAMM: Okay, Steven?

DR. PANTILAT: I agree about the concern about the ratio. It seems like there are many ways to get to an efficient ratio that don't maintain quality. That seems like a real concern. And I think the point that's being made about the cost of achieving quality.

And then I think there's also the sense of the cost of maintaining quality. And so efficiency is an important way of thinking about it.

I was thinking in my world of palliative care there are teams for the same staffing that might see 200 patients a year, 400 or 800 patients a year. And whether the teams that see 800 a year for the same staffing provide quality is actually a really important question. And what it

takes to maintain a level of quality for a certain amount of resource allocation.

And there's sort of another way of looking at it which is what is the resource that you'd need to maintain the quality. And the measure that sort of tries to put these together or at least sees them side by side is actually incredibly important, I think very helpful to the field.

And then just the other point. This came up earlier just about in kind of a moral or ethical dimension to this. But to the sense that you can achieve the same quality at lower cost there is kind of an ethical imperative in that.

If you could achieve same quality at lower cost it does allow you to put that cost -- to either save the money and spend it somewhere else either for healthcare or something else. So there is a -- I think that's something we haven't really talked about but there is a dimension of that that

is important.

And so understanding efficiency that way does have -- it's not just sort of a business thing in a bad way but actually in a very good way, that if you could be more efficient it's actually better for society really.

CO-CHAIR FLAMM: Thank you. Jack?

DR. NEEDLEMAN: I think Andy's comments about composite measures is the relevant place for me to raise the three and a half or three measure thing.

When I look at the descriptions of the methods and what's being done here there are basically only three ways that cost and quality are being measured or pulled together. And they all -- the first one is some variation of the side-by-side. So we keep talking about that as a unique thing.

But in essence in the side-by-side the conditional, the hurdle models, the unconditional models, we measure cost

separately however we're going to do that. And we measure quality or outcomes separately, however we're going to do that. And then the question is how we put them together to display them.

So in the side-by-side we give you your cost score, we give somebody the quality score. This is a five-star plan on quality, it's a three-star plan on cost. Okay, so that's the side-by-side.

Or we can graph it up there. We've got a cost dimension, we've got a quality dimension. And then the different payers begin saying where are we making cut points to highlight specific folks here.

On the value payment where there's a plus payment we highlight the high performers. In the hurdle models we highlight the low performers and we get them out of our model.

But it's the same data. The cost calculated in one way, the quality

calculated separately and then displayed together and considered together. So, we've got -- those are basically the same measurement models. It's just the way they're being used that are slightly different.

The unconditional model is a hybrid. That's why I said three and a half. Because there we make some explicit weight between cost and quality. We add them up and we get a single score.

And if I heard one clear message from yesterday it was don't do that. We don't trust that. We don't trust that we can get the weights right. We don't think the information is there to help anybody make informed decisions.

The two other methods that are here, the regression model and -- actually there are three methods here. The regression model, the DEA which is based on an operations research kind of model of drawing

the frontier, and the stochastic frontier which is a regression-based model for drawing the frontier. So we've got three separate ways of doing the analysis.

Our efforts to graph the cost-quality tradeoff differently and have a very different metric. And when Timbie and Normand looked at a regression model versus side-by-side or some variation of it they said we came to different rankings of folks. So that ought to give us a little bit of pause about just using the univariate models in each dimension.

But that's been, you know, when I look at the list of who's doing what -- with the exception of our friends in the VA virtually nobody is using anything other than some variation of the side-by-side, separately measuring cost, separately measuring quality with all the weaknesses and limitations of how each of those measures are constructed. And we can

evaluate each of them separately.

But that looks like the state of the art. We've got measure each one separately and display them or consider them together in various ways with different kinds of cut points. Or do some more exotic method like -- exotic. Sorry, that's a terrible adjective.

Use some other method like regression, or stochastic regression, or frontier regression, or something else and integrate the data differently.

And we've heard a lot of skepticism here about the integration models. I would characterize them mostly at this point in terms of use in this space as not ready for prime-time even though I like them a lot as an analytic framework.

But I would encourage you to think about simplifying the way you describe the models in terms -- because you mention well, this always starts out with the same two

steps in the unconditional model. That was your reference point as opposed to the side-by-side.

I think it's important to note those two steps are both the measurement analytic -- both of those are the measurement steps. And everything else is just how do we display it, and how do we think about it, and where do we put the cut points.

And I'd encourage you to simplify the presentation here to capture that in a way of thinking about how many different ways do we have to actually analyze and integrate this data into some measure of efficiency or resource use.

And what the efficiency measure is measuring and what it doesn't capture in terms of the quality levels that we want because that's the other thing that's come out of this conversation.

CO-CHAIR FLAMM: Thank you. Steven.

DR. ASCH: I would just like to make

the point that we seem to have come to a consensus that side-by-side is the easiest for people to understand but it's going to be different I think for different stakeholders.

There are some stakeholders that want a more sophisticated ratio and can look at all the quadrants, for instance, and might want a stochastic frontier model. And I think -- I wouldn't want to exclude those kind of analyses.

And then with regard to the waste issue I agree with my esteemed colleague Dr. Almenoff that waste is an incredibly politically popular way of describing efficiency.

But the way I view it is it's kind of the mirror image of the quality hurdle approach. Right? You take the lowest possible quality, meaning either it's harmful or it has no benefit whatsoever, and then you look at the variation in cost. And

any variation above the optimum is waste.

And so, I don't want us to -- even though it sounds like I might be in the minority, I don't want us to go too far and say that the only thing that we can do is side-by-side. I think there are plenty of instances in which the ratio might be useful to sophisticated audiences like policymakers and purchasers.

CO-CHAIR FLAMM: I'm going to let Larry go.

MR. BECKER: Thank you. So, it's interesting because we know that costs differ and we know that ways to achieve costs differ. And I think the same is true about quality. There are different ways to achieve quality. And who's to say what's good enough quality.

And all of that is to say that in statistics you have upper and lower control limits. And that we ought to create corridors.

But I think what's really important and right to the previous point it seems to me that we should come out with a set of stated principles and a stated set of methodologies underlying how, for example, you might combine these.

So if you're not going to do side-by-side here's a series of either principles or methodologies under which you should be combining these so that when other people look at it they can understand it and we don't have a wild frontier out there.

CO-CHAIR DUBOW: So, that's exactly where I'm going. I think that the issue -- I think Jeff actually has been saying this all along. We need to be able to show what we're doing. We don't want to obscure it.

And in the jargon, the contemporary jargon, the buzzword in Washington these days is transparency. I think that there is absolutely an element of trust that is lacking basically.

And I think differently about the Medicare program, for example, and the commercial sector. In Medicare I know that they're going to weight HCAHPS, I think it's 30 percent because we advocated for 40 because we knew they wouldn't go there. So -- well, that's the game you play. The hospitals push back really hard and we knew what we would get.

But it's public. And we know what that weight is for those particular -- for that particular patient experience component.

In the commercial sector when an insurer tiers consumers don't have a clue about what goes into the composition or the creation of that network. And they are -- they don't trust it. They think they're being put in there because it's cheap.

The discussion around narrow networks happened not because they know they're bad but because they assumed they

were bad because they were done for very specific reasons. Exactly what you're talking about, Steven. Because we need to be able to use our resources efficiently to be able to make healthcare affordable.

So I think that the methods that we choose have to be -- and the principles. We have to go to the principles because NQF is not yet engaged in implementation. NQF doesn't tell anybody how to implement something.

We can make a suggestion about pairing particular measures. So sometimes a side-by-side would happen because you say pair this quality measure with this cost measure. But NQF is not in the implementation space. And it relies on employers and insurers and whoever else is publishing these data for the public.

I think one critical principle is complete transparency about what goes into the summation, to the summary score.

Because they're doing it now. We want to know what's behind it. Just say so and then people can make judgments. It's the drill-down stuff. Most people probably won't even look at it. But the fact that this stuff is available will be reassuring or it will direct decisions in particular ways.

CO-CHAIR FLAMM: Thank you. Cathy.

DR. MACLEAN: Just echoing some of the other comments. I think that the very useful output of this really is in defining those principles and the framework.

And Andy, you said something about if we were going to be linking quality measures and cost measures. And that kind of struck me a little bit because I guess I wasn't really thinking about this in terms of cost measures. I was thinking about it in terms of cost. You know, you just measure the cost rather than it necessarily being an endorsed measure. But there are cost measures.

But I think that in this framework the existing cost measures have very limited applicability and what we're talking about is pretty broad. So I think that it would be useful to define what -- how we ought to go about measuring cost. What's the episode. What should be -- that the principle is that you need to define the episode and you need to define what the things are that are in it.

And the same thing on the quality side, that the quality component of this -- we talked yesterday about outcomes. We also talked about that in some instances maybe outcomes aren't the best quality. There's some debate in the room about whether you think you could substitute process for outcome.

But I think what we're looking for is -- or what would be valuable is to lay out that framework. So this is the stuff that should be in for quality. This is the

stuff that we have to think about for cost.

CO-CHAIR FLAMM: Matt.

DR. ROUSCULP: To follow up on what was said. You know, the one group, the U.S. Preventive Services Task Force looked at cost effectiveness. And they put together the gold book and that was kind of the way to move forward in really thinking about bringing cost and outcomes or quality is where they got to.

One area that I would hope that we really kind of winnow down is we're really looking at the precision of quality. NQF has done a great job of really trying to help to identify what are the appropriate measures for quality.

We can also be very precise in trying to estimate what our costs are, be it episodes of care, be it -- however we want to go and measure it.

I think it's important for us to be quite precise on the relationship. So if we

have quality measures that someone is going to come by and say this is what we want measured we've got to make sure that those relate to the appropriate cost. Because we can very easily be very imprecise. We can say this is what a quality of care for heart transplant looks like.

But the cost element I'm bringing in is everything within that hospital, be it whether it's related to the heart transplant or anything else. We've got to be very careful in that because very easily we can get to the point that any ratio that we create can be gamed immensely. And we want to make sure that we are precise and that's the guidance that we give.

And maybe you had already thought of this, but I just want to make sure that we don't miss that because that's one of the big lessons that we learned many years ago when we tried to put together the task force and they're in the process of redoing that

now. And I just think there's some lessons and perhaps some benefits you want to look at.

CO-CHAIR FLAMM: Christine.

DR. RYAN: I think this is an important issue that I think it would be good to get the feeling of the committee on.

Because just think about what's happening on the hospital side, that -- so quality is being measured for heart failure, pneumonia, heart attack, infection. So it's some subset of all the care that's provided. The outcomes are for heart attack, heart failure, pneumonia, right. And then some complication things.

And then the cost is currently being assessed for everyone. So, right now the measurement system in Medicare for hospital value-based purchasing would be inconsistent with what Matthew just said we should do.

And I think a lot of what we've seen in our scan, there isn't a clear -- this is

kind of what I tried to allude to yesterday, that there aren't, you know, the conditions on the cost side and those on the quality side aren't necessarily the same. There's not a whole lot of standardization.

And so I think the hard line would be to say programs just shouldn't be designed that way, that these are principles of standardization in terms of the conditions on both sides or the other elements of standardization that Greg has been alluding to. Or we could just say these are things that developers should consider.

So anyway, those are maybe two sides of how the paper could kind of touch on these. But I just wanted to get a flavor for if what Matthew said is the kind of like belief of the committee that the measures shouldn't be inconsistent in that way.

CO-CHAIR FLAMM: Christine.

DR. GOESCHEL: Thank you. A couple

of things. Many things strike me.

The first is that I think it's really going to be useful and critically important to make sure that the high-level framework for the paper is crisp and takes us where we need to go.

And a couple of things that I heard this morning made me feel better about what happened yesterday. Because on my way home I wasn't feeling so good about that.

The thing that struck me today is the notion that everyone needs to be involved in the space. The recognition that the decisions that organizations or entities make may have impacts on all society, that notion of the pinwheel.

The phrase "rigor and respect" struck me. I wrote it down. Because I think wherever we come we have to acknowledge that we are at the beginning of a messy space. And the minute the white paper is out we're going to know more about

either different methods or different impacts. And what we don't want is people to believe that this is an endpoint rather than a beginning.

Jack, you have an amazing way of simplifying for people like me that are not in the space in terms of framing I'm going to say where we are now, where we need to go and potentially what might be involved in getting us there.

As someone who lives primarily in the quality space I think it is -- I do head jerks because we're talking as though measuring quality we know how to do. And I would argue that we don't.

And so as we think about efficiency and effectiveness and price and cost and how to link all this to quality we're linking to another messy space.

I guess the other thing that -- and it ties into what you said, Joyce, and what I hear, in terms of communicating whatever

we do in a place that engenders trust.

I think I mentioned yesterday that I live in Maryland and we have a new waiver. And when I got home last night I had emails from people in my organization saying you're our new person for quality. We have all these effing ratios that came to us from Maryland. We can't make any sense out of this.

And this is finance calling the quality person. And when we talk to the rate-setting commission they say we use really sophisticated regression models that you wouldn't understand, but trust us.

(Laughter)

DR. GOESCHEL: Because we apply them across the whole state.

I mean, and I think the reality is that organizations and entities are living this every day.

So the notion, and I look at the brilliance in this room, I feel it. And we

get so much into the granular and the potential. I think it's critically important not to stick with just the side-by-side or the variations, to really point where we need to go to have better understanding of what we're doing. But to encapsulate it in a way that doesn't end up sounding like we've -- and I'm not suggesting that side-by-side is dumbed down, but that we don't settle for something and that we don't leave the impression of trust us. Because I think that would be a really horrific and false choice.

CO-CHAIR FLAMM: Kimberly.

DR. RASK: I'd really like to just echo several of these remarks in thinking about the framework.

As I listen to what Jack said one of the thoughts for framework is really to kind of think of this as we have our cost measures, we have our quality measures and we have the question of whether or not you

combine them. And one piece of the white paper is talking about the side-by-sides not mathematically combined cost-quality measures.

We can -- based upon work that NQF has already done about what principles are for quality measures we can speak to that. Any measure that's proposed, be it side-by-side, hurdle, whatever that information, what are the issues with the quality measure, what are the principles of the quality measures, what are the principles with the cost measures, and what are the principles, what would be new, what are the principles with using a hurdle model versus a side-by-side that we would want to address.

Second, there are efficiency measures that are going to try and mathematically combine quality and cost, and that's a new field. There's some information out there. There's not a lot.

What are the principles that a developer would have to be able to demonstrate to have an organization like NQF be ready to endorse something? Or perhaps in looking at the field and what's there to say this is a really promising future area but it's not where we are yet and therefore NQF at this point is not going to be endorsing those measures but encourages the development and study and evaluation of innovative ways to do this. And welcomes work by others who are in this area to help move that forward. Doesn't mean we don't want to go there, but it is kind of a little bit different evaluation strategy depending on what kind of measure it is and it gives us the opportunity to provide some more grounded principles around the one model versus where we may hope to be in the future.

CO-CHAIR DUBOW: I meant to say this before but Andy's remark triggered this.

Because this is already being done, your example of Medicare, I think the principals have to have some kind of a statement about having to express the limitations of the approach so that people understand how you can and cannot interpret the data.

It's happening anyway. So there needs to be some acknowledgment that Medicare is using cost data that may not correspond to the time we measured the quality, for example. Or you know, whatever. But people need help in understanding.

Again, it's not to say that every consumer is going to look at this. But there ought to be full disclosure of how that number can be interpreted.

I think though that what the paper needs to help us understand is how far each of these methodologies can go without really going too far, without really doing damage

in terms of the ability to represent a performance of efficiency. To really distort what we're talking about.

I think we have to be able to say if you do X it's -- you can't. Don't do that. So again, full disclosure of the limitations. How far can we go and get away with it and then where are the barriers.

And I say that strictly from a pragmatic perspective. This is happening and we're not going to stop insurance companies from tiering. In fact, we're encouraging them to create these high-performance networks.

So I think we just need to be realistic about the fact that it's -- we need to give guidance that's very, very practical.

DR. RYAN: And Joyce, can you give an example of going too far?

CO-CHAIR DUBOW: No, I need you to tell me that. Because I don't know. I

don't know enough about the methodology to know how far. But you guys, you guys all do, how far -- where you really distort, or where you don't really say anything that's meaningful or fair. Maybe you can help me, Jeff, I don't know.

DR. SILBER: I think we've given one example that people seem to agree on that the populations should be similar when looking at the cost side and the quality side.

And if you're going to be looking at a combined measure we've also said that you have to -- one guideline would be that an efficiency measure should not be used in isolation to the disaggregated cost and quality measure. In other words, one of the NQF -- a guideline could be you can't present one without the other.

And then we said in terms of when you go too far is when the data for cost isn't related to the data for quality.

So those are just some examples. But we could think of others, other principles.

CO-CHAIR FLAMM: I just want to make one brief comment on this notion of matching the clinical design, the population scope of the cost and the quality side.

To the extent that some of the programs today do look at certain specialty episodes of care and match the episode cost to the quality side, that kind of works.

But as we start looking at measures, and understanding that performance on cost may vary by this specialty, may look very efficient but overall perhaps they're not efficient.

We have this measure of total cost of care that is emerging. We don't have a total quality of care measure. And we're always going to have some degree of imperfect kind of alignment. So getting good enough with good intentions but not

necessarily perfection is something I just wanted to throw in there.

Gary.

DR. YOUNG: Just to echo the points about outlining a set of principles. I don't know if I would frame it so much about it going too far, but I certainly think that an important outcome from the paper would be to outline how these approaches match up against certain scientific principles.

Some of that I think would require empirical work. So yesterday in our own group we talked about the importance of robustness and that we'd all be pretty squeamish if we saw providers being ranked very differently depending on which approach you're applying.

And some of these approaches may lead to more consistent results than others. Some of these approaches may produce similar rankings. Other approaches may produce very different rankings. And so I think those

are some of the principles we'd like to see outlined in the paper.

And some of that may require some empirical work where we can apply some data and see how rankings match up across the different approaches, look at reliability, look at validity, the kinds of central scientific principles that we all live by. I think that's going to be an important consideration in the paper.

CO-CHAIR FLAMM: Jeff.

DR. SILBER: Twenty-five years ago when cost-effectiveness was relatively new and people were bringing that into clinical trials there was a lot of concern about, again, taking data that wasn't from the trial to extrapolate and make a cost effectiveness statement.

So rules were set down. There were task forces that came out from the government -- people probably remember this -- which said, for example, you have to use

the cost and the outcomes from the trial. At least, there might be other things you do, but one of your analyses that you have to report is that they go together. So, it wouldn't be the first time if we were suggesting such things.

And going back to the rules that were set up for clinical trials and cost effectiveness might be one to look at.

CO-CHAIR FLAMM: Okay.

MS. WILBON: So, we are a little off agenda but that's okay because we're making adjustments.

I did want to -- I think this discussion has been great, by the way. We are I think really kind of getting at what we were hoping to get out of this two days and I think the group is on a roll. So we don't want to stop if people still have things to -- that they want to express.

But I did want to kind of bounce it back to Andy and Chris to see if -- I know

Chris kind of jump-started the discussion off with some of his thoughts and to see whether or not the discussion after that kind of addressed some of the concerns and some of the concerns that Andy raised, or whether or not there are any kind of major topics that you feel like need more discussion. And we can figure out how to continue to facilitate that discussion. Or just to kind of check the pulse on what you guys are thinking.

DR. RYAN: The comments I want to make I just made them after some comments in the room were made. So I think my kind of overview kind of direction comment, I don't have anything further to say on that.

You know, I've raised a couple of questions. One about composites on the quality side, what kind of guidance we might want to offer on that that I think is unanswered.

I'm getting, you know, we're getting

some feedback on the notion of standardization. And I think it seems like the committee is leaning towards at least stating a preference for measuring cost and quality for the same populations.

But -- so that's good. But I don't know if you, Chris, want to move the conversation in any particular way right now?

DR. TOMPKINS: No, I think that we're good with this session and onto the next.

MS. WILBON: Okay, so let's go ahead and take about a 13-minute break until quarter of 11. And then we'll come back and move onto the next agenda item. Thanks, everyone.

(Whereupon, the foregoing matter went off the record at 10:31 a.m. and went back on the record at 10:48 a.m.)

MR. AMIN: Okay, is everybody ready for the next section of the discussion?

Please find your way back to your seats. Want to make sure we get everybody out on time.

So, the next portion of the discussion is really to think about the question of what NQF should be endorsing. This sort of cuts to the very operational guidance that we're looking for from this committee.

So for those of you that are maybe a little less aware, NQF has a formal consensus development process which we call the CDP to review and endorse performance measures. We have a long history of endorsing quality measures. We've been over the last three years looking at cost of care measures and many of you around the table have been involved with those efforts. And many of you from the beginning.

The proposed cost of care and resource use measures are evaluated across four overall criteria. The first criteria

looks at the importance to measure and report. The second looks at the scientific acceptability of the measure properties which is really an evaluation of the reliability and validity testing of the measure to ensure that the costs of care that are represented in the measure score represent valid assessments of resource utilization.

We also look at the feasibility of the measure and then the use and usability of the measure to ensure that the results can be used for both accountability and performance improvement objectives.

We've gone through -- so that's the overall criteria that's looked at for cost of care measures. I would just note one of the major questions that's to be answered that we'd like the group to discuss is that currently there's no requirement for cost of care measures that are -- cost or resource use measures that are submitted to NQF to

have any sort of pairing or quality signal that's associated with that, either at the -- well, I'll just say at the measure level for the purposes of this discussion.

So there's still I think some room for discussion around the composite question that Andy raised during our last session. So I'll just point out that NQF's current gc related to composites, I'll just recap a little bit of what we discussed yesterday for the sake of having all the information for the discussion is that the composite measures that are submitted to NQF, and we have a number of them that are endorsed, none of them which include -- I can say this with full confidence -- that none of them include both the cost and quality signal in a single measure score itself.

So the guidance that we have for composite measures is essentially the quality construct itself should be considered essentially. The individual

components of the composite measure should demonstrate a gap in performance. And there should be a conceptual and analytic justification for including the components in the overall composite measure.

And each of the components of the composites should provide added value to the overall composite, either empirically or conceptually. And that reliability and validity of the overall constructed composite measure should be demonstrated. So that's a little bit of just the general framing of how the current guidance related to composite measures again which currently the only composites we have include components that are of quality measures only. So there's still I think a little bit of discussion to be had around what potential guidance we would expect if you were to include multiple different types of measures.

So the overall questions that we

have for the committee are what should NQF be endorsing in terms of quote unquote efficiency measures. Should it be the programmatic methodology of the approach, a combination of measures, or both?

How might the current endorsement process for cost and resource use measures integrate linking quality measures? Specifically, should resource use measures be evaluated in the context of identified quality measures that would be linked or evaluate them together, noting that that would be a significant departure from where we are today.

If NQF were to evaluate efficiency models -- and I think we've moved away from that potentially, but that's also an area for discussion -- for endorsement which criteria or principles should be considered in terms of testing thresholds, feasibility, or any of the guidance that exists in the current NQF composite guidance.

And I think the last one is probably a little bit out of scope considering the discussions that we've had over the last day but I'll just for the sake of completion just ask it. What are the implications for endorsement of efficiency measure/models that include rating systems, i.e., the Star rating system.

MS. WILBON: I would probably slightly re-frame that question maybe in the context of some of the discussions that have come up potentially around the side-by-side.

And I think a place that NQF has not gone that Joyce mentioned before is kind of in the implementation space. To date we don't really evaluate the way that the measures are technically displayed. So we look at kind of the construction of the measure, the actual testing and so forth that goes into the development and construction of the measure, but not necessarily how it's displayed.

So I think this whole notion of the side-by-side display of potentially the efficiency measure profile if you will, that that would be somewhat of an added piece to kind of our current process in terms of just looking at measures individually.

So, just based on the discussion I think it would be really good to hear some feedback on whether or not that -- we need to be looking at how people might be proposing to -- or entities or organizations might be proposing to display that information, whether that should be part of the endorsement process as well. So.

MR. AMIN: So, yes, I would just summarize this discussion that we're hoping to have before lunch to say that currently NQF looks at endorsement of cost measures and endorsement of quality measures. And that's the extent of what we're doing.

What should be in addition to that, if anything? I think that's inherently the

question being asked. And we can -- there are various different levels of what can be done from there. But at a very simple level that's the question that's being asked. Joyce?

CO-CHAIR DUBOW: Taroon, if you can just remind me as a policy matter hasn't NQF taken a position -- I just seem to have participated in so many conversations where we advocate for providing information -- when you provide information on cost that you should also provide some information on quality. Have we no policy position at NQF that says that?

MR. AMIN: So it's operational guidance. I would say the way that we sort of, you know, that it's not clearly in the criteria. There's no requirement in the criteria but it's guidance as -- yes.

And as we've gone through the governance structures of NQF, the standing committees, the Consensus Standards Approval

Committee and the board in particular, all those levels have reaffirmed that these measures should be used in the context of quality. I mean, there's more layers to that but effectively that's the -- where it stopped.

I mean, there's no specific requirements as measures are submitted for endorsement that they're linked with quality measures or definitely not anything at the programmatic level at this point.

So I would ask -- I mean the series of questions is basically should there be something more than that and then maybe a further clarification around how would -- what would the pathway be to get there? So any clarification?

Use your microphone. Thank you.

DR. YOUNG: I just want to clarify this because I think I heard yesterday that there's been a change in policy at NQF that composites now don't require endorsement for

all of the individual measures that comprise the composite. I know that was the policy in the past but now it's the composite itself that is endorsed and the individual measures that comprise it don't each individually have to be endorsed by NQF.

MR. AMIN: That is correct.

DR. YOUNG: Okay.

MR. AMIN: That is correct.

CO-CHAIR DUBOW: So are we looking for discussion on these questions? Okay. So, you all can see the questions in terms of what NQF should be endorsing in terms of efficiency measures. Do we all understand what we mean by that? Tim, you want to take a leap?

DR. LOWE: Well, I don't know about that. I was thinking though in terms of -- I understand it's kind of new ground in terms of providing guidance.

But I do think that how these are -- the end product does obviously influence the

selection. So, if we're going to select a methodology that has implications for display then obviously that has to be specified.

Now, it may just be guidance. Maybe you can't endorse a particular look. I do think that the field comes up with some very creative ways sometimes of explaining things.

I know I give it to colleagues and I'm somewhat color-blind so they always do a better job than I am. And some of the stuff is just gorgeous, what they create. So there's creativity out there.

But if they know what the basic structure needs to be, you know, because I think people get lost in this. But also I think without that end goal I don't know how you would design the process to get there.

So I don't think I can answer all these questions for you at the moment.

CO-CHAIR DUBOW: But we did have

that conversation yesterday about what the purpose of the measurement was. And you know, it seems to me that this is a very, very broad question, but that we need to be a little bit more granular in giving some guidance. Because I think it all depends what you're using it for. So. Greg.

DR. WOZNIAK: I guess I'm not sure if you're kind of taking these dot point by dot point, but at least a couple of comments. One on the first one.

And I think you just said it. There needs to be the flexibility to allow for the developer and the user to have the kinds of methods or use the kind of methods with the kind of data that they have available.

And one of the things that's apparent in the discussion, a lot of these techniques, these models, they're driven by what data is available, what you can do, what you can't do. So it needs to have that flexibility.

But jumping down to the last point. I guess I'm unclear, are you talking about how the data is visualized, or are you actually talking about things like thresholds? Because there was a lot of discussion yesterday about we should not be considering anybody, or we should not be giving pay-for-performance bonuses or whatever to anybody that's in that low-quality tier.

I mean are we talking about thresholds of values for some value of cost and quality, or are we just talking about we're using Stars, or we're using the Consumer Report little circles. I always forget the names of those. Like STS uses those. Harvey balls, okay.

CO-CHAIR DUBOW: Is it Hardy or Harvey?

DR. MACLEAN: Harvey. Half reds and whatever they are.

DR. SILBER: We call them blobs.

DR. WOZNIAK: Whatever, right. But I mean, are we talking about pure visualization, or are we talking -- which then has sort of the -- what's the information.

CO-CHAIR DUBOW: Yes, I think that needs some clarification.

DR. WOZNIAK: Or is it thresholds we're talking about? I would avoid suggesting thresholds because NQF doesn't say performance measures, you need to meet a threshold of 60 percent or 70 percent. But which are we talking about?

MS. WILBON: I think we were talking more in the -- my re-framing of that last bullet was more in the display, more in the visualization and whether or not that should be integrated as part of the evaluation given the kind of --

DR. WOZNIAK: I'd probably say no there too.

CO-CHAIR DUBOW: Actually, I don't

understand what that means. I mean, visualization reflects something that is underneath. So the Five Stars and the Medicare Advantage program, for example, reflects a methodology.

It speaks to who's qualified to earn bonuses because the requirements are very particular. The cut points are established so there are methods that are behind it. You know, simply saying -- we're talking about the display, are you saying use Stars as opposed to Harvey balls? Not, I don't think.

MS. WILBON: No. I think I'm just trying to get at the fact that there's been the discussion around the importance of having a side-by-side model to show what the efficiency profile is.

And so showing the efficiency profile in a side-by-side model is key to showing information that kind of visualization of the cost and quality next

to each other is not typically part of the way we ask for the information from them. They may add it but it's not a part of the evaluation process per se. So I guess that's what I'm trying to tease out if that makes sense.

CO-CHAIR DUBOW: So with that I would ask Jeff this question because he's the side-by-side man.

(Laughter)

DR. SILBER: I'm not the only one here. But I think -- I would have thought a scatter plot would have been -- a scatter plot or its equivalent in terms of some information that conveys the same information in the scatter plot and leave it.

If people want to use the different letters -- but I mean -- the point is there's got to be an X and a Y. I think that's the key point.

CO-CHAIR DUBOW: I think the

question I have for you, though, scatter plots will never fly on Medicare Compare.

DR. SILBER: So what I'm saying is there's got to be an X and a Y. I think that's the point. We want to disaggregate cost from quality. That's the thing that would need to be conveyed besides the efficiency measure.

CO-CHAIR DUBOW: When we talk about display in that last bullet, because I really think we should focus on the other questions, but the question is whether this side-by-side has to be right there next to the Stars, or whether you can click down and see, okay, here's what went into the Stars.

Here's what -- there is no efficiency measure in the Star rating for Medicare Advantage right now. It's strictly -- it's some administrative stuff which I think is problematic but there's no efficiency assessment. It's HEDIS measures, it's CAHPS and it's some -- it's how fast

you answer your telephone for Part D.

DR. SILBER: I guess the point is -- we wouldn't want it to be that what we require is for an efficiency measure that somehow the details of that measure are in terms of its functional form. People won't understand that.

But they will understand that the parts of the measure in terms of cost and the parts of the measure in terms of quality laid out in some side-by-side fashion, that will understand that tradeoff that we're talking about.

So it's -- when I heard earlier today, we're talking about what needs to be described and transparency, I agree you do want to be transparent about whatever efficiency measure you make.

But one important element of the transparency is to give the two dimensions. So if it's done by a scatter plot, or it's done by a blob for quality and a blob for

cost and they're put next to each other, or a number for quality and a number for cost and they're put next to each other and you have an element of the distribution across the different choices that you can make in terms of other hospitals that you would go to or providers and this is where they stand.

I think it's the disaggregation of the two axes that are needed.

DR. PANTILAT: I just find myself in this discussion wondering if it might be helpful to think about a specific example rather than a general example. And if there was a way to look at a particular quality measure and then think about what the -- how you might think about a cost measure that goes with it and then try to answer these questions including how we might display it.

Just thinking about it in theory is a little bit more -- for me at least is being a little more challenging than

thinking about some of the measures that were presented yesterday, some of those composite measures and others.

If you chose those as measures of quality how might those be combined with cost to be an efficiency measure?

CO-CHAIR DUBOW: I think we shouldn't focus on the display itself, but rather what's behind the display. Because I don't think this group is going to be the one to be the dispositive voice on display. That's a whole different science.

DR. PANTILAT: And I don't even mean to focus on that, but just how might you combine it with cost in a way that -- because the quality is very specific to a population, to a measure. And then how might cost be equally specific that way to combine it into any kind of efficiency measure.

CO-CHAIR DUBOW: -- an example from the hospital side. I think you really are

conversant in what's happening in that space right now.

DR. RYAN: So, with hospital value-based purchasing it's kind of unique. And Chris is more the expert because he was one of the architects of this.

But it's -- each of the individual measures kind of get filtered through this attainment and improvement like paradigm where attainment is basically based on these standard benchmarks. You get these points for attainment and then the year over year changes, you get points for improvement. And then each measure gets a performance score.

And then these scores then kind of go up to the domain level. And then at the domain level they get weighted and put together as a total performance score.

And so with hospital value-based purchasing there's this efficiency domain that's I guess online. Is it next year? I

think it's next year.

DR. RASK: It's now. The Medicare spending per beneficiary? It's out.

DR. RYAN: And so that gets weighted, I don't know, 30 percent, something.

And so the domains get weighted, you get a total performance score and then there's a payment adjustment based on that.

But, so -- and this is kind of the unconditional model is based on our definitions.

Now, what some private payers would do for cost is they -- some did a statistical model where they said, okay, like these people are -- let's take episode cost. These people are significantly different from mean cost so we're going to put them in kind of high- and low-cost kind of categories and then everybody else is in the middle.

Other payers were just kind of

classified based on the distributions like quartiles or something based on cost. And then you'd have like, you know, maybe four categories.

And then what was common for the tiers is to say, you know, if you're, you know, if you're just some combination of that distribution. Like if you're just bottom quartile for cost, high quartile for quality, we'll say you qualify for distinction.

And so that would be a way that the measures would be operationally combined is typically there's some kind of categorization classification of the cost input into some box. And then similar on the quality side and that's how they're put together.

CO-CHAIR DUBOW: Does that help you think about the question a little bit better?

MS. WILBON: So we're just

displaying an example if this helps from a measure that we've received in the resource use measure evaluation process on how they're proposing to display the cost measure in the context of the quality measure. Thank you, Jack, for sending this to us.

But I'll just add, and I feel like I probably derailed the conversation or confused it by re-framing that question. But we didn't endorse their approach to doing this. We just endorsed the cost measure. So, the question for the group is really more about --

CO-CHAIR DUBOW: What was the measure?

MS. WILBON: So the measure is the cost measure around, I don't remember what this one was. Is this the total cost? I don't know the measure number.

(Laughter)

MS. WILBON: But again, this was how

they were trying to show the committee how they're using the measure in the context of quality. So this was their -- okay, thank you. So I'm just showing this as an example of someone was saying that they would like to kind of see an example of --

CO-CHAIR DUBOW: What's the unit of analysis?

DR. NEEDLEMAN: This is the per-member per-month -- fee-for-service participants in Medicare per member per month attributed to the physician group measure. The practice measure.

MR. BECKER: What's the quality?

DR. NEEDLEMAN: Oh, we weren't asked about that one. So I don't even remember what they were displaying here. I can find it out. I've got all the documentation here. Give me a moment.

MS. WILBON: But this again was not intended for you guys to go into detail. It's just to show you an example of what we

have been presented with but we haven't really gone into an evaluation on the details of the questions that you're asking. We didn't go there because that's not currently part of the process.

So again, I feel like I'm taking us down a rabbit hole going further. But again, it was just purely for illustration purposes to kind of demonstrate where we are right now in terms of our evaluation process.

We were given a cost measure and said to evaluate under a separate -- under current criteria and said -- and they told us we are going to be using this in the context of quality and this is how we're going to use it.

But we didn't go down this rabbit hole of, well, what are the quality metrics. How are they going to be combined. How are you determining, you know, what the scatter plot numbers are. We didn't go there

because we don't have the guidance to do that.

CO-CHAIR DUBOW: So was this intended for public display? Was it intended for regulation to look at outliers? What was it intended for?

MS. WILBON: This is a CMS measure. So it would have been provided to the physician groups.

CO-CHAIR DUBOW: It looks like a CMS measure.

MS. WILBON: It would have been provided to the physician groups to tell them where they are in the feedback report.

DR. PANTILAT: This is to evaluate a specific cost measure? A way of assessing cost?

MS. WILBON: Cost and quality.

DR. PANTILAT: And quality. No, I understand that. But there's a specific cost methodology here and a specific quality measure that's here.

MS. WILBON: Right. But we didn't evaluate the quality measures, or what quality measures they were using. It was only the cost measures.

DR. MACLEAN: We're kind of having this little side conversation related looking at the specs for the CMS physician value-based modifier program which does this.

And I just kind of wonder if it would be instructive for the committee to kind of go through this. I don't know how many people are familiar with it.

But I think the thing that strikes me about this program is it's very nicely laid out. And they've got tiers and so on.

But when you look at the quality measures which are basically the PQRS measures it's kind of a whole bunch of different measures. But then the cost measures are pretty, you know, limited to a couple of conditions.

And I think that's something we ought to have some discussion around and maybe make some recommendations on whether there should be some sort of relationship. If NQF is going to be looking at cost measures if there should also be some sort of a quality measure that goes along with it.

You know, there are different -- like the -- I was also kind of pulling up the -- Joyce, you probably were involved in this. The patient disclosure program. The consumer purchaser disclosure thing a few years ago which basically a bunch of people signed onto it and plans agreed that yes, if we were going to report out anything about cost on docs we also have to report it out on quality. But it didn't get so specific to say that it had to be on the same cost and quality. Anyway.

CO-CHAIR DUBOW: But this does raise the question that Greg talked about before

in terms of the flexibility that he's looking for in terms of data sources and availability of data. Remember, Congress mandated the value-based modifier and CMS had to hustle and come up with something.

And I wonder whether there are challenges with respect to data. Certainly there are limitations in terms of what measures are around to do it within the time frame. Which is a whole other story that needs to be recognized.

We live in -- but I think the question of data availability is an issue. And how much flexibility NQF wants to be able to exercise in terms of looking at some of this stuff.

I mean, do we take into account some of that stuff? Are we agnostic about what the challenges are in developing these things?

DR. BURSTIN: Just a few thoughts. I think these are all good questions.

I think there's always a tradeoff between specificity and comparability. You just have to put it on the table. You can give flexibility and then you're left with measures that are not comparable.

So I think particularly on the cost side these are primarily claims data. Availability isn't really much of an issue. It's more of an issue I think, my sense of it, being the quality person more on the quality side of very different data sources producing very different results. Here it's not quite that way.

I do think these are, you know, we got a lot of heat to be honest when we endorsed cost measures. A huge amount of heat. I mean, endless comments, a couple of people were on that committee saying how could you. You said as part of a prior framework report that NQF did not believe in looking at cost measures in isolation. It should always be done with quality. And

here you are endorsing cost and resource use measures.

And that was the slide that Taroon showed earlier. That building block slide was exactly produced in response to that to say yes, but to get to efficiency we have to have reliable valid measures of cost and resource use to combine with quality to get to efficiency.

So I think we're still at the same place now. We've now got some measures we feel pretty comfortable are cost and resource use measures. I agree with the comments earlier that we've got a long way to go on quality too in terms of which quality measures you would actually put with these cost measures.

Some of the PQRS process measures don't seem like the kinds of things you would throw in a box with a total cost of care measure, for example.

But I think, you know, the question

I think really for all of you is should NQF take a different approach. If in fact, and we talked about this early on the cost measure comes forward should we ask the developer identify what is the related quality measure. Whether it's side-by-side, combined, whatever the case may be, but it should be related to.

And you see that in the value-based purchasing program from CMS where they've at least got here's the cost measure, here's the outcome measures. But those are exactly the kind of questions we'd like to hear from you.

And again, you know, the endorsement criteria are intentionally fluid. We update them every year. If there is again needed updates to this around implementation and how to look at these and perhaps put more teeth around what has just been guidance we are really looking to all of you to provide that information to us.

MR. BECKER: So, as everybody knows consensus is getting harder and harder as we get deeper into the cost issues, the reimbursement issues and all of that.

And it seems to me that I think everybody in the room would agree that as we've gone through this subject there doesn't appear at least to be a right answer. And there might be a consensus around how we want to go, but we don't know all the intended and unintended consequences of whatever it is that we ultimately recommend.

So as I think about processes like these in a Black Belt process you would run a pilot. I wonder if one of the things we should recommend is some rapid cycle learning events around let's figure out or let's recommend that CMS, for example, run some of these recommendations in small pilots with how you display this kind of data this way, or this kind of data. And

let's learn. Let's do some rapid cycle learning. Because I don't know that if we put 100 more learned people in this room we'd come to an answer. So I think there's a lot that needs to be understood about this.

And I think if we write this paper and we recommend that we run some processes. And then thinking about when we're trying to get to consensus and we're trying to actually get to adopt a series of criteria lots of good thoughts but I think we're going to need some hard evidence that we've really looked at this and here's what happens when.

CO-CHAIR DUBOW: Let's just start down there with Gary and come up the line, please.

DR. YOUNG: I guess as I think about this it seems to be there's sort of a truth in advertising feature to this. Because payers, purchasers want your endorsement

because it gives their measures or their approach a legitimacy with respect to providers.

Clearly you can endorse a cost measure or some other measure, several measures, and then they can put it together in ways that in fact don't necessarily meet the criteria in which the individual measures were initially endorsed for.

And so what are they saying to their provider community? Are they saying, you know, are they saying the individual measures were endorsed? That a provider may come back and say yes, but you're holding me accountable to something that's different. You're holding me accountable to this matrix, or this complex set of measures, or this composite. Has that been endorsed? Well, I think the response to that would have to be no.

If they want endorsement for that then yes, it has to go through NQF. So I

think there's sort of a truth in advertising feature to this.

Payers, purchasers want your endorsement because that provides a sense of legitimacy that they can then communicate to their providers. This has been endorsed.

But if you're combining different measures that have each individually been endorsed. So the example you gave us before, the cost measure was endorsed. You didn't ask questions about how they were going to use it. Well, I think that's fine.

But if they're going to use it in some sort of a process that didn't get endorsed then they can't communicate that to their providers as being endorsed. If they want that endorsed then they have to go through the NQF process.

CO-CHAIR DUBOW: Steven?

DR. ASCH: That's what I was going to say.

DR. YOUNG: Well, we discussed it,

so it's all right.

CO-CHAIR DUBOW: Jack.

DR. NEEDLEMAN: The conversation has moved on. Anybody who wants to know what was on the quality side there I've got it up. Come see me at lunch.

I'm coming back to the questions here. And I've sort of migrated down to the bottom of this list of four items rather than the top.

I think use matters here even though NQF does not endorse for use. And the reality here is people are being on these measures. People are being tiered on these measures. People are being excluded from these networks on these measures. So there are sharp cut points in how the measures are being used.

Most of the measures are like that scatter plot we saw, two different measures. So, each of the measures needs to meet all the standards for endorsement. If we're

talking about composites each of those dimensions, the X and the Y as Jeff has said I think to me need to be endorsed.

The issue that's come up here which is not as the criteria were explained when we were on the Cost and Resource Use Committee, that is not fully there in I think the current standards that should be is the concept of robustness.

We've got the split sample of things to show validity. If we do the same thing but we do it to two slightly different data sets or two subsets of the data we get the same answer.

But what we don't have answered here, given the uncertainties about the data is whether if we use slightly different weighting systems, a slightly different risk adjustment system, if we include the drug data or don't include the drug data because we only have it half the time do we wind up with the same rankings.

And I would encourage that as we think about this given the high -- given the nature of the way in which these are being used that the robustness of the measures needs to have more consideration. And that needs to be incorporated into the conversation as well. So that to me would be one of the things when I look at 3 and 4.

NQF, and indeed when we were looking at the cost measures for CMS we were not asked about where the cut point should be.

But if you're looking at a Star system the cut points very much matter and they're very much part of the measurement system. So we don't have criteria for cut points. We don't have criteria for when are differences material and when aren't they. We've got these sharp lines being drawn.

But if everybody is about the same quality drawing a line that distinguishes the top 10 percent from everybody else isn't really capturing anything material. So the

distributions matter and that also has not been considered when we've been looking at these things. So how much variance there actually is matters.

NCQA says they take that into account when they're saying do we want to go down this path from creating a measure in this area. Are we ready to retire this measure because everybody is basically at the same level.

So NCQA some of whose measures we've considered and endorsed takes that into account. The question is whether the NQF process should be taking the degree of variance into account as well related to this issue of cut points.

So, those are some of the things that strike me as I think about the ability to endorse here. Each dimension needing to be endorsed, the robustness, thinking about cut points and use and the implications of that for how much variance and what the

variance is telling us.

CO-CHAIR DUBOW: You mentioned two things. The issue of robustness I think is an arguable question about whether that is included in the criteria. I would suggest that it is implicitly included in the evidence requirement, isn't it? And it's a judgment. Isn't that -- you don't think so?

MR. AMIN: I think the issue that Jack is bringing up is around the current criteria related to validity. We only -- the NQF criteria allows a systematic evaluation of face validity which doesn't get to the whole issue about the stability of rankings that I think Jack was pointing out.

CO-CHAIR DUBOW: So that brings me to the second point. And I wonder whether having a clear statement of principles where we can articulate the areas where we need to be robust and how we could be robust as well as to address the second point that you made

would advance this conversation about cut points and things of that sort.

Can we come up with a set of principles that address some of these concerns in a way that -- I mean one of them you talked about is, you say, use matters. And we have had that theme throughout these two days. We agree that use matters.

So, is it helpful to think -- one of the bullets here does ask us to think about principles, to be able to identify some principles that would incorporate some of these ideas.

MR. AMIN: Can I also tag onto that, Joyce? I think the point that you're making is a really important one.

I think there's two points that have been raised here by Jack and the second one I think Andy has raised earlier.

I think the big question that, Jack, you're raising is purely tied to the whole question of use. You know, how these cut

points are, how the stability of the rankings, depending on what the purpose of the -- the way that the measure is used drives how much the requirements need to be in terms of some of the validity testing.

And again, I just would point out I welcome discussion on this topic as well. I mean, NQF has had its position right now and I think we're certainly rethinking it, at least more explicitly, about whether or not the criteria need to be maybe not different, but at least the way that the criteria are applied need to be clearly -- clearly need to be potentially different depending on the use case.

And when we say use case that means, you know, we broadly define accountability applications as public reporting and pay-for-performance applications. And you could imagine there's a gradient there even.

So currently we don't think about those as necessarily measure properties.

Those are things that happen after the measure is endorsed. And you know, it's a reliable, valid measure. It's good enough to be used for payment. It's good enough to be used for public reporting. And there's really no distinction currently. I think that's open for discussion.

And some of these questions around cut points and the robustness I think really relate to once you're starting to use it for payment or limited networks the requirements may be potentially greater. And I would submit that for discussion.

I think the second question that, Jack, you just raised and Andy just raised before is the whole question of the profile. When you actually look at this graph, you know, Andy, you raised I think earlier today or yesterday the question that you might actually need to do reliability and validity testing of some sort of this actual profile.

That's not something that we've ever

done in the past and I'm not -- I mean, I analytically -- I'd be curious to understand how analytically that would be done. And if we could provide some additional guidance on whether that's something we should be requiring and how it should be, you know, what we would expect in terms of a submission from developers, that type of clarity would be welcome as well.

So, I think the two that have been brought up here that I would welcome some additional conversation around is this question about the profile and what exactly we would require in terms of the profile.

And then secondly, this question around use case and how much that should be driving some of the criteria here.

DR. SILBER: I agree. I was thinking about what Jack said. Usually we're not in the context of having two variables, it's one variable. So I don't want to make a joke about double robustness

which is not anything to do with this.

But we're doubly concerned. And so I think Jack's point is a really important one.

Again, going back to the cost-effectiveness literature and the cost-benefit literature people didn't get too hung up or haven't generally been about needing to be robust for the cost and robust for the benefit. But they have worked out proper confidence intervals that are those ellipses that deal with the cost and the benefit. So, I mean, some -- and of course some of the papers that were referenced dealt with this.

But a statement from -- in this document saying that when you're doing this even with the side-by-side it probably isn't a bad idea to have some confidence interval around those points. But that's a separate issue than the robustness.

But I guess I'm saying I totally

agree with what Jack said. It's double important because of the new problem that we're looking at. And there is some literature from the past that could be drawn on in terms of cost-effectiveness.

DR. SCHUUR: So I would echo the point about use being very important here. I think the value-based payment modifier metric is a good example to look at which is a continuous cost -- continuous adjusted cost versus a performance on quality.

But really it's a tertile of quality versus a tertile of cost. And so when evaluating it I would suggest that the NQF criteria should really look at that.

Is there evidence to support the cut points for quality. Is there evidence to support the cut points for cost. Is there validity of those cut points.

And then the reliability test should be testing those cut points as opposed to testing whether or not the model on costs is

reliable in a linear or a logarithmic framework for a continuous cost variable. Because the important thing is going to be how providers get ranked for those.

DR. MACLEAN: Just one comment on that. So I think that there are times where maybe there wouldn't be a statistically significant difference between the different tertiles or however you break it up.

But I would just caution that it's still important to know the cost and the quality and the efficiency. Because maybe there are some areas where we as a society or as a country don't do so well. And I wouldn't want to lose the opportunity to understand those point estimates to try to improve it.

DR. SCHUUR: Let me just clarify. I agree with that. And I think the case for cost can be made. You know, cost is important and so variation in cost is important.

But rather than saying -- if your model is looking at is there reliability on the measure on average cost and you compare using traditional reliability metrics you're comparing the dollar costs.

What's really important is are providers, groups, hospitals going into fall into the high, low, or medium groups. And what proportion of them are going to be reclassified differently under the different model. Because that's how the measure is -- that's the effect of the measure.

And you may get a very -- the actual -- that's really the characteristic of the measure is in tertiles. It's not being used as a continuous cost model.

CO-CHAIR DUBOW: Back to use again. Jack?

DR. NEEDLEMAN: As I've been sitting here listening to the conversation I had two additional thoughts. Again, partly driven by use, partly thinking about principles and

criteria here.

One of the things that's come up over and over again in the last couple of days is this concept of you can be -- you can efficiently produce low-quality care. And I have not heard anybody say that's acceptable.

So, the efficiency measures, however you measure them, even if you're measuring something different from cost, really do need to be coupled explicitly. And even if you're computing them with DEA or regression or anything else, they do need to be coupled with some sense of what the quality is. So that to me is one of the principles I would want to see in terms of endorsing a pure efficiency measure.

The other thing is I do think the conversation is different when there's evidence -- evidence that there are many low-cost high-efficiency high-quality providers, that that is an attainable state

given the current technology and the current way we organize and deliver care.

Because to me the issue of payment, the issue of tiering, all are much easier if I know there's no tradeoff between cost and quality. So I think it's important for those who are trying to think about using these measures, think about whether that space is filled, or whether that space is empty. Because if that space is empty there really is a tradeoff between cost and quality.

And the conversation about what we're trying to do in healthcare, where we make the decision about what the minimum acceptable quality is, whether we think there's a maximum acceptable quality because achieving anything more than that is too expensive is part of the conversation. But that's a different conversation than we can have our cake and eat it. We can have high-quality care at low cost.

So, looking at what the nature of the relationship is between cost and quality, whether efficient providers, whether low-cost providers produce high-quality care is -- or the tradeoff is to me part of the discussion of use and part of how we think about the issues -- the decisions we have to make using these measures. And those two conversations differ depending upon whether that low-cost high-quality quadrant is fillable.

DR. SCANLON: Jack, just -- I mean, are you holding price constant in that, or no? What's your view on that?

DR. NEEDLEMAN: No. Well, what I'm saying is if we map, you know, go back to that thing. If we map cost against quality, however we measure cost, however we measure quality, and that whatever -- that sweet spot box of high-quality low-cost has folks. So that's feasible.

If I can get CLABSI rate down to

zero without spending any more money then I'm going to insist everybody go to zero.

But if the only way that I can get to that highest quality is by spending more, that's a different conversation, Dennis. That's do we want to pay the amount to bring everybody up to that level because it's going to cost. Is there a lower level of quality that we think is the sweet spot and we're not going to pay to move everybody up to that higher level.

That's a different conversation about the -- the efficiency level to achieve high quality is not a low-cost level. That's a different conversation about the cost-quality tradeoff than if we don't.

And I think in terms of moving the quality and the efficiency agenda forward we have to know which decision space we're in.

The frontier can be very squared in which case we don't have the tradeoffs because the sweet spot is there. But if the

frontier is not squared then we've got the tradeoff and we need to be able to talk about that in ways that don't get us into death panels.

(Laughter)

CO-CHAIR DUBOW: Okay, Jeff.

DR. SILBER: It seems to me that I'm concerned about an insurer that would want to save money by using tertiles rather than quintiles. So that they can group the quality of the very best along with the little bit better than average in the top tertile.

And therefore -- and if they're not going to provide the continuous measure then the use of the tertile can be used in a way that's devious.

I don't want that to happen. So maybe what should happen is we might not have to make a statement about whether you have to use tertiles or quintiles or quartiles or whatever. But we could say

whatever you give you also have to give the continuous measure. Therefore, we could ensure against the elimination of the understanding of really good quality which is beyond the upper -- it's not just that you're in the upper tertile, right?

So I guess -- I don't want the game the system by playing with tertiles or quintiles. But I want the continuous number as well. And I think that will be another aspect of what we put into the report.

CO-CHAIR DUBOW: Greg?

DR. WOZNIAK: I think yesterday I said my condolences to Chris and Andy. Today you have my sincere condolences.

(Laughter)

DR. WOZNIAK: I think we got back to a lot of the discussion has been around one of my comments about are we talking about thresholds or are we talking about visualizations. And it seems like a lot of this discussion has been around thresholds

and whether or not we should be talking about the thresholds or the properties of the thresholds. And whether or not we should be trying to improve care and improve quality. I think we all agree to that. I'm not sure how that goes into this report or how this feeds into the report.

The details around this get really, really messy as we've heard. I mean, there was the RAND report that did the reliability study on ETGs and they showed the reliability depended on the specialty that you were looking at. And obviously sample sizes. So now are we going to start saying about you could use this, or this is reliable and robust for this set of specialties but not that set of specialties?

So I'm a little bit concerned about how far we go and how much detail there is, and that it be consistent with some of the other NQF frameworks in terms of what are the sort of high-level criteria that you've

applied to other types of measures.

So again, I don't think -- I go back to my comment. None of those endorsements and none of those criteria say things like performance on a measure ought to be 70 percent. So I'm not clear that we can say something about the threshold for minimum quality needs to be X. And how you actually then have the measure developers use that?

Because what's going to happen is they're going to stop going to NQF to get endorsement because there's all kinds of measures that are being applied outside of CMS, for example, and even within CMS that don't have NQF endorsement.

There's all kinds of private payers that use measures that haven't seen the light of day of NQF. They're all home baked, they're all homegrown.

So you've got to be concerned about not only -- my sympathy to the authors here in terms of what they include and don't

include, but what you put out and how far you want to go in terms of your criteria and what you want to actually consider or not when these measures come to NQF.

CO-CHAIR DUBOW: But what about the principle around robustness? Can't you apply to different scenarios the principle of robustness?

DR. WOZNIAK: Well, again, I think you can. It depends on how far you want to go and how far reliability, robustness, validity all tie together. Because those things all do tie together and there's different metrics of those. And then it's a question of what metrics would be acceptable as a robustness criteria or a reliability criteria.

DR. BURSTIN: Taroon hinted at this earlier, that we've had a lot of discussions about whether the idea of a binary yes/no endorsement has outlived its usefulness, and whether it's time to move to endorsement

that is perhaps more fit for the intended purpose.

So you might, for example, just playing this out. We don't have internal consistency I think on staff this is a great idea, but we know we need to do something other than the yes/no.

But might you, for example, incorporate that robustness when you know the intended use is for high-stakes financial payment as opposed to something where it's perhaps a lesser --

MR. BECKER: As opposed to high-stakes lives.

DR. BURSTIN: No, no, no, I was not going to say that, Larry. I'm saying high-stakes financial as an example. I didn't say high-stakes lives. I'm just saying as an example as opposed to something that might be used for benchmarking or QI.

DR. MACLEAN: So I think our discussion today is really kind of

highlighting some of the many challenges in this. And I think that Andy and Chris, if the paper could lay out what we hope to gain out of this sort of a system and what we want to avoid. And I think then kind of work into what the way you do it would be helpful.

In some of the discussion this morning, particularly with the concern about if you have a low efficiency you're going to have high efficiency, low-quality care. So I think that argues for having a threshold.

But yesterday we -- there were concerns expressed about if you have a threshold then maybe that would tamp down trying to raise up scores even higher. And Jeff has now raised a concern about tertiles, or quartiles, or however you put them together, that it's going to be less transparent.

My concern about laying out the point estimate is that you probably can't

validly discriminate against -- the confidence intervals are going to be so wide at the point estimate that I think that there is methodological value to putting people into groupings, or entities into groupings that you can discriminate against. So, that's another consideration.

But I think we want to just lay out all the concerns and then maybe back into what the model ought to be.

DR. ASCH: This is the problem with putting your tent up and then having seven people ahead of you. However, so I'm just going to agree with two different things that people have already said.

And that is if there are going to be threshold measures then the continuous measures should be made I believe transparent or even published or attached, despite the problem of confidence intervals.

I think the threshold measures will have the effect that they're going to have

whether the continuous measures are attached to them or not. And for all the reasons that people have already said so I won't go into them again.

Then, Helen, I had thought for a long time that NQF should stop the yes/no thing and start talking about how the measures are going to be used and what the optimum use case for a measurement is.

And that there should be a few paragraphs associated with every measure saying this is what we at NQF think the ideal use of this measure is. And if people choose to use it differently as they probably will then they do. But at least NQF has put its flag in the ground.

And other agencies have done things like that as I'm sure you know at least to some extent.

Because really measures aren't just measures. Measures are how they're used. And that's why we keep coming back to the

use case. And rather than be silent on it, why don't we at least say what we mean.

CO-CHAIR DUBOW: Okay. Staff? Where are we on this?

MR. AMIN: It appears that we've exhausted the discussion. Is there more? Joe, it seems like you have -- no?

Yes, I mean there are a number of committee members who have been part of this discussion from the standing committee's perspective. And so we'd welcome the committee members' discussion or thoughts on this topic.

CO-CHAIR DUBOW: I think that's a really good question, how helpful this would be to you all on the standing committee.

DR. NEEDLEMAN: Well, not just me but also Andy who's actually writing the paper and a few other folks around the room. Okay. So I think this is extremely helpful.

DR. WOZNIAK: I'm sorry, Jack? Sorry. Could you maybe explain to the rest

of us what the standing committee is? I guess we're not exactly clear.

MR. AMIN: So, across all of the various different consensus standards -- across the various different clinical areas and topical areas that are looking at measures we have standing committees that are tasked with two to three years of being a sitting committee with the purpose of owning a portfolio in a particular topical area.

So, we've seated a Cost and Resource Use Standing Committee. Many of those members are part of this panel as well. Jack, Joe, Herb, Andy, there are others. Larry, obviously. So a number of folks that are around the table. And so this input is really to be given back to the standing committee in terms of how they should be evaluating cost of care measures going forward.

And this group was particularly

tasked with the question -- this group meaning the linking cost and quality group was tasked with the question of whether we should be expanding anything related to our criteria for cost of care measures. And that's been a topic of debate with the standing committee for quite some time now, probably last year.

So again, I'd welcome thoughts or reflections from the standing committee members on the topic that's been the discussion for the last hour.

CO-CHAIR DUBOW: These used to be called steering committees. They have changed their name. They're a little bit different now.

But the reason they're important at NQF is that they make recommendations. And those recommendations go out for public comment.

And NQF receives the public comment. The committee reconsiders or considers the

comments, reconsiders its recommendations and then the recommendation goes forward. And then it goes to the CSAC which stands for the Consensus Standards Approval --

(Laughter)

CO-CHAIR DUBOW: So it goes to a committee that again consider it but looks at it from a policy recommendation.

The bottom line here is that the recommendations of the standing committee are very important. And the purpose of having principles and guidance is to inform the thinking of the people who sit on these various committees, to help them figure out how to apply some good -- some principles to their deliberation so that they reflect NQF policy. So it's very important work.

So, okay. And I apologize for putting you on the spot, Jack. I forgot that there were others here.

But it would be useful to know whether we are heading in a right direction

and whether there's more that you need and what you think. Because I think we're probably heading for lunch soon, but we need to be sure that both Chris and Andy have what they need again, that we're getting closer to having a set of useful principles so that we can draw this conversation to a close with respect to how NQF is going to actually use this stuff.

MR. STEPHANSKY: I have been very encouraged, particularly today because yesterday I think we were still kind of wandering around in the dark.

And I have a feeling that there are things that we have not really articulated here that are going to be in the paper that will be even more useful.

I think the issue of the yes/no is also going to be important. And I hope we can continue to work on that one.

While I have not had much to say to the group we've certainly been having all

sorts of conversations in the background that had -- just are starting to really become rich. So I'm very encouraged.

DR. NEEDLEMAN: When I think about the committee deliberations that we've had, and we've had from what I've heard from staff some of the most contentious, lowest consensus votes on the measures.

And I think it reflects the state of the criteria and the sense of use as much as the nature of the measures that have been brought forward.

The issues that have sort of emerged out of those discussions are very much in the quality of the measure. So things like the completeness of the data. The exclusion of often drug costs sometimes, behavioral health costs because of carve-outs have driven some of us absolutely crazy in terms of judging whether we've got complete enough resource use measures.

The fact that we're working with

billing data has all kinds of other issues related to what are we really measuring here and what resources are visible and what resources that are used in care are invisible.

The standardization of pricing. All these very technical issues.

The issue of Bayesian shrinkage has not emerged very much in our discussions but probably ought to come back into those.

But a lot of technical issues about what data is counted, what's counted, what's not have been issues.

We haven't discussed that very much over the last couple of days but I think those will continue to be issues.

Some of the issues about robustness and the relative rankings we have discussed in ways that we sort of snuck it into the usability conversation. And snuck it into an extension of the reliability conversations beyond the traditional ways of

measuring reliability of measures.

And we have -- some people on the committee have been more aggressive in doing that and others have been more reticent to do that.

I think the conversation we've had today sort of reinforces that and in that sense has been very helpful.

We've been treating them as cost measures in some ways, relative resource use measures. And the conversations we've had in the last two days around efficiency versus cost and value versus cost that should be fully reflected in the paper I think are going to be helpful in reshaping some of the conversations and clarifying what's being measured and what's not being measured. What's being excluded here.

And as people try to do more efficiency measurement, composite measurement around efficiency I think those -- the conversations we've had are going to

be -- and the paper are going to be extremely helpful in informing the committee and helping the committee reflect on what we're trying to do, what the limits are, what we're being asked to look at. So those are some quick reflections. I'll probably have a few more as we continue the conversation.

DR. WONG: So, I won't belabor the point, but I share the same view that Joe and Jack have just mentioned.

I think the high-level remark that I have is that I think that over the last two days the conversation where we're blending in the cost component and the quality component together has helped put a better framework, I guess that's the way we say it, put a better framework in terms of how these different components are functioning. So I think that it is incredibly valuable.

DR. RYAN: What I would say from the perspective of the Cost and Resource Use

Committee is that these have been really complicated discussions. There's been a lot of issues, maybe kind of alluded discussions in some sense. But I think good issues that people need to work through.

I think if we added a dimension to that that was also simultaneously evaluating quality measure and then the linking in that same structure it seems like it would just be overkill. It would be information overload and be very difficult to come to consensus.

And I think if that committee knew that there was a process in place to kind of link cost and quality and we could feel kind of good about that, that NQF is saying, okay, this is how you should do it. Then I think that would make, you know, that would make the job of that committee a lot easier to kind of just focus on the issue at hand or on the cost measures. That's the issue.

And knowing that there's a guidance

or there's a process or there's, you know, I'm not sure how NQF wants to go with this, but potentially kind of an approval process that we buy into to combine these so we can just kind of focus on the matter at hand.

CO-CHAIR DUBOW: A lot of pressure on this paper.

MR. AMIN: Okay. So I think we have one more comment from Matthew.

DR. ROUSCULP: No, no, that's okay. Just for clarification. And I know it feels like we've been going around this a little bit which is about what NQF is seeking out of this.

I mean does NQF want to be able to say look, we want to have a way to assess efficiency measure or measures to make sure it goes forward, or is it more of saying, look, we're creating these cost measures, we have quality measures that we are approving, and if you're going to look at efficiency, you're going to use these, these are things

to consider and to think about going forward and allow that to just be your stopping point.

Because with both of those that kind of dictates where your paper would go or what your ongoing work will want to think about. Because I think it's a big leap forward for you if you want to get to that point of being -- giving recommendations.

MR. AMIN: I think you certainly identified the question.

I mean, in a lot of ways NQF doesn't have an objective in its own right. The purpose of what we're here to do is to reflect the consensus of the community on what's needed.

So, the basic place that we've been is that there was a need for quality measures, there was an increasing need for cost measures. We heard feedback that if you're going to be using these you should be using them in the context of quality.

The question really is how far should that go. Should the evaluation just lie on the fact that of, okay, as we're going through cost measurement, you know, there's an acknowledgment that it should be used in the context of quality. And that could be where we stop.

Or is the fact that the measures are being used together in these types of potential applications, that for particular applications that it would need to be -- the evaluation would need to be more robust to the point of evaluating the methodology of how these are brought together, the actual profiles are evaluated more systematically.

Whatever application that you could imagine. There's obviously the four that we looked at today. They might have a different threshold for what, you know, how we look at these two things together.

And ultimately the question that we're trying to answer here is how far

should that go. Should it be more than what we have now which is just sort of an acknowledgment that they need to be looked at together? Or does it need to be more robust than that.

And I think what I've heard is that depending on the use case it may need to be more robust. We need to consider some of these other -- you know, we can't overwhelm the system in terms of requiring specific cutoffs and things like that.

But also resting on the fact of if you're going to be putting these two signals together the methodology needs to be much more transparent than it is now. So that's at least some of what I'm hearing here. But again, if there's other thoughts around that topic.

DR. ROUSCULP: Because if you're still at the point where you're saying look, you're hoping that the white paper will help you think through this a little bit more I'm

wondering if at the very end if you almost have to say look, if NQF wants to get into this point of being able to assess, kind of say yes, this is an appropriate efficiency measures, that's going to require a whole lot of additional work that goes in one direction.

And is that -- at the very end it says look, here's the options for NQF at the very end and here's all the work that would be associated with each of those options.

MR. AMIN: I'm not hearing a specific consensus around one. So I would imagine we're going to be on the latter which is going to be here are a few options and here's what's going to be required to get there. I think that's generally what I'm hearing. But again, I'm not -- Chris, I'd welcome --

DR. TOMPKINS: Well, yes, I think that's probably where you are going to end up.

What I've been struggling with is that sometimes these -- some of these models lend themselves to being more in your turf. Because if somebody says you know what? I'm going to produce an efficiency measure and therefore in the submission requirements they have to do all the diligence of saying.

For example, just drawing from the literature that's been mentioned so much. The cost for AMI. Cost for AMI discharge, you know, mortality rates is the quality measure and the cost and so forth, like that.

If you're going to say, all right, I'm going to model this. I'm going to model this. I'm going to squeeze all the juice I can out of this with a cost-effectiveness measure and something like that. Then that brings it into your measurement framework a rubric already.

And so -- but going back to a lot of this conversation started at least in my

mind back when Gary said about the difference between someone claiming that they're using NQF-endorsed measures and then using them in a particular way that he described as a matric which I think corresponds to other people's use with a scatter plot. And the using it in the particular way is where the real action is.

And if NQF was to say, okay, all these models are good, for people wanting endorsement of the efficiency measure they bring it all to you. But if they're trying to do the efficiency they can still play off tradeoff of your endorsement by using the scatter plot side-by-side method but just doing it in any way they really wish to configure.

So the gray area question for you is how much, and the ensuing conversation around that discussed that a bit, how much can you reach out or push your commentary on how these ought to be used. Just a slightly

different -- I think we're about to break.

Steve actually stopped me in my tracks a long time ago. He didn't mean to. With your question, the way you framed it which I thought was good which is we ought to think if you have a given quality measure in mind what should be the linkage, right? What should be the logical linkage of the methods.

It reminded me when I was sitting in that corner with the QI group for awhile yesterday when people were saying things like just because you're efficient in this doesn't mean you're efficient in that. We might be the best in town for this and we might be the worst in town for that.

So really what's running all this is that it's the cost of what, quality for what, and then linking them together. And the number of instances in which that plays out.

Some people know that one of my

current activities is building a grouper for CMS. So at this time next year we may be unleashing on NQF and everybody else 1,000 resource use measures.

Because, you know, you ask the question is somebody efficient in managing osteoarthritis patients. That's one question. Which is -- and the efficiency might be the rate at which they use medical supervision and other things to avoid surgery and conservative practice and so forth. You might be very efficient at doing that.

Separately is the question what about the osteoarthritis patients who actually have to have major surgery. Given that that's true or that happens how efficient is that?

So the way you, you know, you're going to end up talking about hundreds of conditions and embedded in those conditions are many different treatment options. And

these are all opportunities to put microscopes or spotlights or flood lamps on different concepts of what efficiency is.

So that's why you stopped me. Because if I'm about to roll out 1,000 -- your logical question is what is the linkage process to make it appropriate to frame each one of those as an efficiency measure, whether it's any of these models, as opposed to just the standalone cost measure.

And just as an advertising plug, in the context of the episode grouper I've argued for the logic and the necessity of a grouper because it has some of the qualities that I've stressed over the last two days which is it makes the determination of how you're assigning cost to particular activities explicit and discoverable rather than implicit.

I don't know how people build resource use measures that are just saying well, let's assemble some of these cost

items and call it that when the fact is that when dollars are spent there are sort of plausible reasons why they could be categorized different ways. And I think having a logical system that does all that creates a discoverable largely objective and debatable process for how that happens.

But now we've got the other dimension on top of it which is how do you make those logical linkages over to the quality so that each of those resource use measures becomes explicitly an attempt at value or efficiency.

And that's the NQF goal here. If we did it all in one measure it comes right to you. But if it's all done after the fact in the bushes, you know, sort of out of sight then the real action -- going back to Gary's point -- the real action is maybe happening outside the context of the community standards or consensus process.

CO-CHAIR DUBOW: Steven, you have

the last word before lunch. And public comment.

DR. PANTILAT: I'll make it a quick word then. But just on that point. You know, again, the example I know best is palliative care.

But if you just ask the question what are the costs of patients seen by a palliative care service it looks like palliative care makes everybody more costly because we see the most costly patients.

And so the way you assess cost has to be very specific to the question that you're asking. And it's the costs that accrue after you see a palliative care patient.

And a lot of people look at this and say oh, palliative care makes everything more costly but that's actually not the case.

And so the way you assess the cost and which costs are considered and when in

the course of care they're considered for the service that's provided is incredibly important to tell whether -- what kind of impact you're having. If you took a global measure it would be -- it would make it look the opposite of what it is.

CO-CHAIR DUBOW: Could we ask the operator for public comment? Operator, is there anybody on the line who wants to make a comment?

OPERATOR: Any committee members who want to ask a question please press *1 on your telephone keypad.

CO-CHAIR DUBOW: Nobody?

OPERATOR: There are no questions.

CO-CHAIR DUBOW: Thank you. Is there anybody in the -- here today? Do you want to make a comment? Okay.

MR. HAIDER: Hi again. Thank you so much for taking public comment.

So sitting from the back of the room here I feel the highly knowledgeable group

has quickly -- you've dived into very complex details.

And I guess I would urge the group to kind of divide -- you know, this is a big problem. So in an engineering world if there's a big problem I'll divide it into small portions and then address kind of low-hanging fruit first, and then say okay, we're going to address the rest later. So it's kind of like address the forest before you start diving into the trees.

So maybe the report has a simple problem, defines a simpler problem, and then offers a multi-stage process, baby steps before you run. So that's one general comment.

The second thing gets into a little bit of the details and that is I pulled up the financials of some hospital groups. All the hospitals publish their financials.

And so specifically Johns Hopkins Hospital made $1.8 billion in patient

revenue with an operating income of $70 billion in 2013.

And so the business world would use this as how profitable are they. But what they do, and I look at a couple of other hospital reports, is they only talk about patient revenue, some other revenues in the income, interest, et cetera, and then they talk about how much they spent in salaries and how much they spent in supplies.

Well, in the financial world when I look at a company I'm evaluating I'm looking, well, how much did you spend on marketing. How much did you spend on R&D. How much did you spend on salaries.

So I think maybe some metrics that just come to mind that maybe would be relevant to an efficiency white paper would be, for example, doctor salaries per revenue, bad debt per revenue, total salaries which includes all the hospital staff per revenue, supplies per revenue.

And then have some benchmarks to say, look, if you're spending this much on doctor for this much revenue that's a benchmark. And how they exactly -- how the percentages fall out would be probably interesting. And then comparing that with the quality metrics that you're discussing. So those are my comments.

CO-CHAIR DUBOW: Okay, we're going to reconvene in a half hour at 12:45. Thank you very much. Very interesting conversation this morning.

(Whereupon, the foregoing matter went off the record at 12:13 p.m. and went back on the record at 12:51 p.m.)

CO-CHAIR FLAMM: Okay, if everybody can start moving towards their seats we'll go ahead and be getting started in just a minute.

Okay, so this next section that we are going to have is really to start to summarize where we are, be summarizing our

recommendations and the path forward for where we're taking the work after these two days of discussion.

We have about 45 minutes or a little close to an hour for this discussion. And in your discussion guides there are several questions that we want to come back and reflect on.

I think we have started to get a better idea of what this is going to look like after today's discussion. But if you could forward to the discussion question slides.

So these are the four questions that were in the original discussion guide as we've been reflecting on them. The first is we can the evaluation of efficiency models be use-agnostic.

Probably -- a lot of discussion has been that use is very important. So, rather than spending a lot of time on that discussion we'll probably focus more on the

lower questions.

Reading those through what are some of the benefits and unintended consequences that might result from the endorsement of efficiency models or efficiency measure composites. And we have talked a good bit about that.

And then particularly the final two bullets. What challenges might present for current and future measure development efforts to align with the recommendations that are coming out of this work and what challenges might present for current and future programs to apply the principles and recommendations coming out of this. So it's from the development and from the user program perspectives. So I'll kind of let those questions sink in a little bit.

While you're thinking I guess one of the things that did come up earlier in the day that I'll mention from a program perspective as you're applying these things,

and we did talk about this earlier, is the issue around data availability. And some of these models, particularly ones involving modeling, regression modeling or frontiers imply that you have all the primary data at the patient granular level.

And given that the quality measure data kind of lives somewhere and rolls up to summary statistics that are reported and then we want to use that. And given all the work that's happening in other venues around alignment of measurement, alignment of reporting and not wanting to create redundancy in that space I think we have to be cognizant that there's this imperfect availability of the kind of data that are needed for some of these combination models depending on your program -- where you stand and what you have access to. And there's not universal accessibility of patient-level data. Which is an understatement.

Comments?

DR. ROUSCULP: I'll jump in.

CO-CHAIR FLAMM: Go ahead, Matt.

DR. ROUSCULP: So I'll tell you everything that's absolutely wrong so everyone else can tell you why, what the actual right answer is.

I will say that for the evaluation of efficient models to be use-agnostic, we'll start there. I think it will go back again to where you're going to go.

If you feel as if NQF should be in the arena of saying yes, we're going to have an NQF-approved efficiency type score or direction I think if you ever go to that you will have to go where you can't have use-agnostic.

You're going to have to go drill down to say in provider-type measure of efficiency, or in care provision. You're going to have to kind of drill down and you cannot be use-agnostic.

If on the other hand you're going to

step back to say, all right, we are not going to make recommendations on efficiency model but we know others are going to jump in and start using cost and quality in some sort of comparison then I think you can say look, here are some general rules or directions or I think it was -- I forget what Jeff had said, principles. And I think at that moment you might be able to go into something that's more agnostic for whatever its use might be intended.

CO-CHAIR FLAMM: Dennis?

DR. SCANLON: I'm not sure if I'll articulate this well and it very well may be wrong also. But what strikes me is we're in an incredible period of incentives for innovation, entrepreneurship and ultimately efficiency.

And there's a lot of this happening sort of between parties in ACO relationships, deciding who to partner with, you know, sort of who's going to be your

best partner and everything else.

And I think from a societal perspective that's a good thing.

I guess I wonder -- and to some degree private data which would be outside of the scope of what we're talking about here. I think there's probably strong interest to make sure that private organizations that are developing those relationships have good information on who's efficient and who provides high quality in part because this has to be marketed to patients around narrow networks or sort of you need consumers to kind of come to your plan.

So I guess I wonder whether there may be, you know, I don't know if it's false positives or false negatives, what the publicly reported data that's endorsed says. And then privately what organizations are doing as they sort of partner for the dance so to speak.

And I'm not sure if that's a problem. I guess it could be if, you know, sort of what's out there publicly maybe somehow lags what the private sort of organizations have and they're trying to kind of sell products.

So they're selling products and trying to justify their narrow networks, saying hey, this is high-quality, this is efficient care, maybe to employers, maybe to others. But if that information somehow is disconnected to sort of what's publicly reported for whatever reason in my mind I just wonder sort of how that may all play out. So I don't know if that makes sense but I find myself sort of trying to reconcile how fast-paced and moving the incentives shared savings models and other things are for entrepreneurship, innovation, efficiency which is a good thing with kind of the public reporting.

CO-CHAIR FLAMM: Iyah.

MR. ROMM: So, I guess I would say, building on that, and this comes a little bit back to the I think yes/no question.

I think that there is -- there needs to be a world in which we I think build evaluation of these models in a way that is both use-agnostic and use-case specific.

That is, I think in some ways there is an opportunity for that side-by-side to have some consensus standard as to what sort of the true north to Larry's perspective on efficiency measures or linking of cost and quality measures is.

But then in the actual implementation and/or endorsement of measures I don't see a world in which we can be use case-agnostic.

I think that one other comment sort of building on this earlier principle of benefits and unintended consequences is I think that the rich conversation that we had this morning around what efficiency is and

how it fits into this broader construct of -- the linkage of cost and quality as its own premise and linking towards value is that as endorsement process is undergone that precision is going to be absolutely critical.

And ensuring that the endorsement mechanism is sensitive to the fact that there are many different perspectives on efficiency and the NQF's may not be the only one. And therefore building out this idea of efficiency towards the end of that linkage is critical.

DR. YOUNG: I'm not entirely sure NQF has to make that decision about whether or not efficiency measures have to be evaluated based on use.

I think again, going back to some of the points that were made earlier, maybe also the point that Chris was making before the break.

I think NQF has to send a very

strong message to its stakeholder community that it cannot leverage the endorsement of an individual measure into a larger evaluation process that then claims it's been endorsed by NQF.

You know, if the FDA approves a particular two different types of medicines or two different molecules manufacturers can't combine them together and say this has been FDA-approved. The combination may be toxic.

Well, the same thing here. Combining multiple measures in some way that may on its own may be reliable and valid, but combining them in some way may in fact be toxic. So that can't be -- the purchaser can't claim that that's been endorsed.

So I think that's what you want to communicate to your stakeholder community, that, you know, they can only claim endorsement for what in fact NQF has endorsed and can't extend it beyond the

boundaries of the endorsement.

CO-CHAIR FLAMM: Jeremiah?

DR. SCHUUR: So, a somewhat similar comment I think. Thinking about challenges and the path forward there's probably agreement amongst everyone that we want to improve efficiency and there's a lot of room to improve efficiency in United States healthcare.

The market's clearly creating measures that they're labeling as efficiency measures. And I think the challenge for NQF is as an organization that's focused on standards evaluation and trying to do that in a reliable somewhat scientific way to make sure that the response to be -- there's a pressure to be in this space to be relevant, to be responsive is not to necessarily endorse everything or feel a pressure to endorse measures in the space that may not meet standards of efficiency.

And so I do think that the use case

is important as has been said before because you want the NQF endorsement to mean something.

And I think it's okay for measures that are being used out in the marketplace to be used. And some of them may have NQF endorsement, some may not. But there may be no measures of a certain type that get NQF-endorsed and that doesn't mean they can't get be used. But just -- if they don't meet the standards.

CO-CHAIR FLAMM: Cathy.

DR. MACLEAN: So, echoing comments. But first of all, the first question, can evaluation of efficiency models be use-agnostic.

We've had a lot of discussion about linking cost and value and then efficient measures kind of separately. And I think that they're different. So I think that I'd like to broaden this up a little bit to be linking cost and quality. I think I said

cost and value. I meant cost and quality.

And efficiency is one. When you actually kind of put it together then you're going to have an efficiency measure.

And with regard to whether they can be use-agnostic I think that there are principles that could be use-agnostic but I'm not sure that there's an individual measure that's necessarily use-agnostic. And I think that we should lay that out.

CO-CHAIR FLAMM: Jack.

DR. NEEDLEMAN: I'm thinking back to the first cost and resource use committee meeting I was at which is relevant because it sort of got us to this committee meeting where people were unhappy just looking at cost measures or resource use measures.

Not just because of all the issues about whether you could do it accurately enough which remain but because as Helen said and I think, Carol, you and Joyce, you said, originally NQF said we don't do cost.

It's all about quality.

And that initial committee was uncomfortable. The efficiency word came up. Efficient, effective use, production of high-quality care was the vision of what we really wanted to be able to say something about. And there was a recognition that the cost and use measures were intermediate to get us there.

And I think to think about how to get us there was what helped create this committee.

And I think the conversations of the last two days plus the conversations on the phone sort of reinforce the importance of doing that.

These measures are being linked in the world. And the need to not just look at cost but to relate it to the performance of the system I think has all been reinforced by the conversations that we've had today.

And in that regard I think one of

the most important things that we've done in terms of the conceptual stuff over the last two days is we've taken the efficiency word and we've really said we're talking about measuring the efficient delivery of high-quality care.

That you can't talk about efficiency alone, you have to talk about the quality of care that's being delivered efficiently or inefficiently at the same time.

And that linkage, that tight linkage to the continued concern that the care be high-quality I think is one of the most useful framing devices for thinking about how these measures get used, what the implications are, how an individual cost measure gets interpreted.

So that I think has been extremely important and really a major -- we've been talking a lot at the conceptual level, one of the major conceptual strengths of the conversation the last couple of days.

Related to that is the concept of use. And what we've been asked to endorse at least on the cost and resource use side has been the measure. How do we score people.

Use is all about the cut points. You know, who's high, who's low, what do we do with that information.

And in addition to getting the measure right, getting the cut points right is an important element of use. And so, can the evaluation be use-agnostic? We can ask whether we can get the scale right, accurate, robust, all the other words we've used.

The question I think is whether NQF committees should be engaged in thinking about what -- looking at the measure, looking at the distribution, looking at the imprecision, precision, what we say about the cut points. And I don't think we've fully resolved that today. But that's one

of the ongoing challenges and questions that relate to being use-agnostic versus use-specific.

I think we've got to -- one other thing that struck me about the conversation, and we've had it in a couple of different ways, is that in some sense the, you know, to the extent we're talking about efficiency as our shorthand for talking about this the cost agenda and the quality agenda are a single agenda.

The goal is to produce high-quality care as efficiently as we can. If we only focus on the cost, or we only focus on the quality we lose where the system has to go. And that becomes part of the rationale for thinking about the linking.

But we're not trying to produce cost-effective -- sorry, we're not trying to produce efficient, low-quality care. It's improving quality, improving -- reducing cost, improving the efficiency of delivering

high-quality care as the rationale for why we're measuring this stuff and how we want it used.

CO-CHAIR FLAMM: Those are helpful, thank you. All right, I have Cathy.

DR. MACLEAN: Jack, following on that, something you just said, you said the cost per unit of high-quality care.

Well that means you have to define a cut point, right? You have to define what threshold means high-quality care. And I do think that that's important to define that absolute threshold.

I just want to call out though that that's not the way CMS is going with some of its programs.

So, the Medicare Advantage Stars program right now, the current way it's set up is that for many of the measures to get a four-star rating there is an absolute threshold. But that's going to go away and it's all going to be relative scoring.

The specs that just came out for the exchange is all percentile-based. So it's all relative scoring.

And I don't think that's a good idea because I think that -- while on the one hand it promotes improvement I think that there does need to be a -- we need to define what is good enough.

And then, you know, above that, absolutely, let's report who's better than good enough, right?

But this is a big issue. And if -- I think it would be instructive to kind of explore that in this paper. This issue of absolute versus relative.

CO-CHAIR FLAMM: Steven, thank you.

DR. PANTILAT: So, in thinking about this issue of benefits and unintended consequences, I mean it seems like it's great to find high-quality low-cost providers. And if you can do that, that's great. But that's not always going to be

the case.

And it seems like the other benefits of linking this is to understand how does, for example, cost reduction impact quality. So if you're looking at how do I reduce cost at what point does that impact quality. And a measure that looks at those two things could look at that. Or how do you reduce cost without changing quality.

And then I think the other piece is that very often better quality is going to cost more. And it gives you a way of understanding that as well. Those seem all very important to understand.

I think the unintended consequences, just getting back to something Jack said which is, you know, you could have an efficiency measure that would say that low-quality care looked great. And that to me is an important unintended consequence that could come out, that you could look great by providing really lousy care but doing it

really, really cheaply and efficiently. And that would be very wrong.

I think the other unintended consequence is that somehow cost and quality would not really be linked. And so you would be sort of looking at a cost measure and a quality measure that really aren't linked.

And so that the costs that you're looking at aren't really closely related to the quality that's being provided. And so that changes in one wouldn't necessarily lead to changes in the other.

Again, just in my own world a lot of people think palliative care should reduce length of stay in the hospital. That may or may not be true. It's almost impossible to prove.

And so if that was the outcome where you said palliative care and length of stay were going to be linked somehow there isn't a lot that you can -- that may be true, it's

just very hard to demonstrate. And so you would not want to link those two as an efficiency measure because it's not going to be possible to demonstrate that.

And you may find that people want to link those in a way that really doesn't make sense.

CO-CHAIR FLAMM: Timothy.

DR. LOWE: I have a little bit of a concern about the cutoffs. I do develop measures and I have to tell you, my focus in measure development is always to -- this is how I was trained -- make sure the science is done as best they can.

So I want to make sure that when we're constructing a measure that it is valid and it is reliable and it has a strong evidence base.

That's one reason why we do publish our measures, why we go through peer review, to make sure that it's the best it can be.

I personally feel that the cutoffs

really are the responsibility of regulatory bodies. And the reason why I believe that is because you can game anything. If I know what the end result is going to be I can design the measure to create that. So to me the two things have to be separate.

Now, I realize that if we are adding quality and cost in an equation that we have to have some way and we have to understand that. But there may be a way of bifurcating, whatever, how we decide to divide it in such a way that we've done it statistically that it's kind of -- that's where the agnostic part of this is.

It has to be outside of the political part of this which always gets involved and also out of the desire to game it. We have to find some way around it. And so that's how I've approached it is to focus on how can I best make the measure and then leave it up to the market or the regulatory body to see how they're going to

apply it.

CO-CHAIR FLAMM: Iyah.

MR. ROMM: So, just back to Cathy's comment. I agree with her on the idea of needing to apply both thresholds and this idea of relativity.

But I would posit that it's definitely going to vary across use cases. And so thinking about that in the absolute in and of itself is probably too narrow.

And so I think that we likely will find ourselves in which relativity is appropriate for certain use cases and not for others.

CO-CHAIR FLAMM: Joseph? Oh, sorry.

MR. STEPHANSKY: Pardon?

CO-CHAIR FLAMM: Go ahead.

MR. STEPHANSKY: I want to throw something in here that I actually disagree with Jack on one thing. And it's the only thing I disagree with him on. And this is also going to drive Joyce a little bit

crazy.

Let's consider hospitals for a moment. That if volumes are really going to go down which is part of what we'd like to see, less utilization of hospitals, the hospitals that are out there, they're going to have to make some choices about their strategies for the future.

Many of them simply have figured out, okay, the only way I'm going to make it is I've got to eat that other hospital's lunch. I've got to get their volume to make up for my reduced volume.

And that's just part of how -- a lot of different hospitals are going to have to think about what their strategies are going to be. Some of the small ones who don't have a lot of physician practices and post-acute care components may really have to think about relatively small niches and kind of stay with fee-for-service contracts.

And one of the -- out of all the

different kinds of strategies that hospitals can adopt we have kind of an interesting one. I won't name the hospital system in Michigan but --

(Laughter)

MR. STEPHANSKY: -- they're going for the low-cost good-quality label. Not the high-quality label, the good-quality label. Because there are employers who want to look for that. It is a market niche.

And I'm just offering that when we keep thinking about thresholds that we have to realize that there are situations where some people -- they might not want to buy a Yugo, but they might want to buy a Chevy Malibu and that's good enough.

So when we start thinking about this we've got to understand that hospitals are adopting strategies that may not match what we think they should be, but they're real strategies.

CO-CHAIR FLAMM: Jeff.

DR. SILBER: Malibu's a nice car.

(Laughter)

DR. SILBER: It's a really, really good car.

Related to that and it's very similar, and it's relating to what Cathy had said who's now gone.

So my worry is if you have a threshold and you meet the threshold, and I'm just saying this again, but I just want to reiterate it.

And I worry that you then might diminish the importance of improving quality to its higher levels. So I really worry that I don't want the introduction of a threshold to eliminate the importance of knowing what the quality is all along the continuum above that minimum threshold.

And I guess that's my biggest concern. The way Cathy had stated it, and I put my card up to react to Cathy, is if, well once you get to that threshold, then

you can look at cost. I might be misinterpreting what she was saying, but -- and then we can all look at efficiency in one analysis.

But I think it's really important to not do that. And that we can't fall into the trap of saying that if you're above the threshold quality is in a sense the same.

Because I think plenty of consumers and insurers really also would like to know about those highest -- the highest quality hospitals. And they certainly might cost more but provide the very best quality. So I wouldn't want to group everyone above the threshold in the same category.

CO-CHAIR FLAMM: Matt?

DR. ROUSCULP: To add to that. So once you do set that threshold it's a cost minimization exercise. That's what efficiency is, it's a production function of cost minimization. And I think that's where the big gap is.

I guess the question that comes back is we have talked about the efficiency kind of being that no man's land that sat between cost and value. And so the question comes by is that the value -- however you get to that value and that value exercise, I think it's going to be very important.

Because efficiency alone as a measure can very quickly get to the issue that Jeff had brought up, unless you have that value element to it.

CO-CHAIR FLAMM: Larry.

MR. BECKER: So I wanted to react to what Joe said about thresholds.

And while I understand the implications of a threshold and while thresholds might be important they're not -- in practice application from my perspective as a designer of a plan that serves 250,000 lives is somewhat impractical.

And the reason is because I've got people in Xerox who perhaps are literally,

maybe not literally in every zip code in the United States.

And so when I go into a market I need -- let's take primary care docs. I need so many primary care docs in my network to be able to cover those people. And in some geography, you know, if you think about these cost and quality metrics somebody might not meet the threshold but I still have to have them because people still have to get care.

And so there's the practical side of setting thresholds that when you try to implement something you might need to go below that threshold. Then what? What does that mean? So anyway, I put that out there as just food for thought.

CO-CHAIR DUBOW: Well, I don't actually understand the concept of a threshold for quality. What does it mean? Does it mean you meet basic competency requirements? Does it mean you're licensed?

Does it mean you -- that you're board certified? Does it mean you have MOC? What's good enough? You know, it's laden with values. Which I think is what you're saying, isn't it.

You know, I think it's very hard for the public to understand that somebody is making me settle for something that I don't quite know what I might be deprived of.

And I think that's a real -- so you know, even though we don't love the relative scales, at least we know that that's kind of where everybody is right now as opposed to what I could really be missing out on if I'm -- I mean, you could really get screwed that way. I don't think people understand it.

So, I'm a little uneasy about these thresholds. Who sets them? I mean, it's just -- in addition to -- Carole and I were talking about that earlier, Larry, exactly the point you make about the very practical operationalizing challenges of doing it

which are not easy ones. Those are societal questions about when do you decide to get rid of a really bad hospital.

You know. I mean I don't know how you design these minimum standards to take care of public health and all of the rest of that stuff.

MR. BECKER: But I think that's what transparency is about.

CO-CHAIR DUBOW: Well, I don't know if transparency is good enough because we do want to have a certain modicum of protection which is the regulatory function. And we do that through licensure.

And that's kind of what that's supposed to be, isn't it? It's the bare -- it's the necessary minimum to get into the game.

CO-CHAIR FLAMM: So, I'm going to go to Peter.

DR. ALMENOFF: So I've been listening quietly for awhile and I keep

hearing this concept of high-quality. If you really want to be very honest I'm not even sure. You're not sure about cut points. I'm not sure about our quality.

We have 10 different evaluation systems for U.S. hospitals. You know, the public ones, U.S. News & World Report and all of them.

And a lot of people are embedded in thinking that's real. And so when you drive by a hospital you see we're the best this, we're the best that and everyone buys into it, and they advertise, et cetera.

But when you look at a world-class university hospital in U.S. News & World Report which will do very well. Then you go into another model and they do awful. Because it's really a matter of how you're deciding you're going to measure quality.

So, you know, I think we need to be a little careful about that whole notion of we want high quality. That's apple pie and

I don't think anyone disagrees.

The other thing I've heard was high cost, high quality go together. And in fact, we find the opposite. I don't remember who said it but spending more money doesn't mean you're going to get better quality.

And generally you get worse quality because you do things you shouldn't be doing, procedures that maybe aren't even necessary. And the outcomes aren't very good.

And I think the U.S. healthcare system is a perfect example of that in that we -- compared to the rest of the world we're sort of like a third world nation when it comes to our outcomes. And we spend enormous amounts of money. So clearly there's a problem with what we're doing.

As I mentioned yesterday we probably, you know, I think the best thing to do is develop a framework. You know, I

think all these techniques, some are good, some are bad. Whether it's side-by-side. You know, I think we had a lot of conversation for two days on that.

But I think it's probably the most valuable for at least the public to see or the providers all the different ways that these things could be done, how to do them.

As a user, whether you're a healthcare organization or a provider. I think that would offer a lot more value than trying to say this group is going to be the one who decides what's considered an NQF-validated metric.

Because quite honestly, you know, as many people said earlier today if I'm in a healthcare agency and I'm doing it a certain way and you don't want to validate my metric I'm still going to do it. So I think we need to be very careful and take baby steps and create a basic framework that people can use.

And there are clearly areas where we shouldn't be using certain techniques for certain areas. And I think we probably can agree on a lot of those. But I think it's best to just give them sort of a framework to work in and then maybe in time as these things become a little more mature be able to sort of develop a little more stringent criteria as we move along.

But to sort of say at the very beginning we're going to sort of create this very narrow framework. And we're talking about cut points. To me we can't honestly even talk about quality because I can -- we do that game all the time. I can give you hospital A is the best or it's the worst depending on what kind of methodology I use and what I'm looking at as my criteria.

So, I'm just -- sort of a sign of caution that we clearly have a problem in the United States. We clearly are spending an enormous amount of money for questionable

return.

We probably need to do something. But at least there's not even a break framework on how to go about doing this. So by doing that you're at least I think pushing the agenda and sort of at least getting the system to maybe look in the right direction and kind of at least maybe go forward in the right areas.

CO-CHAIR FLAMM: Jack.

DR. SILBER: Thanks.

CO-CHAIR FLAMM: I apologize. I have you on the list but I did say Jack.

DR. NEEDLEMAN: Okay. We're not going to resolve these issues today.

(Laughter)

DR. NEEDLEMAN: I think it's important to understand where some of the heterogeneity of the vision comes from. Because we're -- to some extent some of this conversation is we're talking past one another when I think we're trying to talk to

one another. Or we think we are.

So there's tremendous heterogeneity in the system. Lots of different services out there. And I think for many of them quality in fact doesn't cost more. Quality saves money. I think that's absolutely right. But it's not clear that that's the case for all services.

I'm going to go back to the P word, the patient word. So what do patients expect of the system?

And for an awful lot of the care that we're talking about patients expect the care if they go into a hospital, they go into a physician, care is going to be safe, the care is going to be delivered reliably, it's going to be to the standard of care that's out there.

And for many of the conditions that are there the result is going to be fairly predictable and only dependent upon the patient's unique physiology and social

circumstances, and not the system that they encounter.

And that's the case for an awful lot of the care we're talking about. I've been studying, Jeff's been studying, Chris has been studying nursing systems for over a decade.

And I think we can talk about J.D. Power's rankings of the hospitals or Consumer Report's rankings of the hospitals on nursing care but the patients don't want that.

I don't think any patient wants to look and say where is the hospital where I'm going to get the best nursing care. They want to simply assume that the care that the nurses deliver is going to be safe, reliable, efficiently delivered, that I don't have to worry about that when I go in.

And for an awful lot of our quality measures it's I don't want to have to worry about that. And in that regard we need to

ask whether everybody can achieve that level at a reasonable cost. And if so we ought to tell everybody, go achieve that level at a reasonable cost. And that's part of what these efficiency measures are about.

For other services there may be real choices that patients and their physicians and their providers face. And Joe's hospital system of good enough may in fact closely match what some patients think.

So if there are hips or knees that we can guarantee you this 90 percent of the time is going to last 15 years but it's going to cost $50,000 to put in, or we can put one in -- and the rheumatologist has left so I can completely make this stuff up -- but there's another device we can put in, it lasts for 10 years in 80 percent of the folks and 15 years in 50 but it only costs $5,000. And you're going to confront a difference of five or ten thousand dollars in your cost-sharing to do this. Do you

want to go the less expensive route and play the odds? Or the more expensive route? And that's a choice.

And that's exactly the kind of choice where Larry's be transparent, be visible, make it clear what the tradeoffs are work. And where we potentially do have this steep gradient of you only get better quality by paying more for it. And people have a right to choose where they are on that grade.

But we need to think about which of those cases we're in measure by measure, care by care.

And we've got ambiguous cases. I don't know whether every patient with rheumatoid arthritis who is a candidate for biologicals ought to be on a biological. Or whether there's such a high cost difference between that and the way we could treat without biologicals that the patient ought to bear some of that cost and bear some of

the decision-making and cost-related decision-making in that.

Right? That's one of these ambiguous cases. Is it the good enough model, or is it the this is the standard of care and everybody is entitled to expect to get it model.

But we need to differentiate as we're thinking about quality, cost, efficiency and the tradeoffs between cost and quality whether we have that, whether we've got examples in that upper left, whether everybody can seek to be in the high-quality/low-cost quadrant where there are real tradeoffs and who should be bearing those tradeoffs, whether they should be socialized in the insurance system or internalized to the patient and the patient's individual decision-making.

CO-CHAIR FLAMM: Okay. Jeff, you've been so patient.

DR. SILBER: Well, allow me to react

to two comments, one from Joyce and one from Peter. I guess I'll start with Peter.

I guess when we -- the whole endeavor that we're in is in some way making an assumption that it's worthwhile to look at these things such that it's -- that we have a decent measure of quality and a decent measure of cost.

I think it's a little bit cynical to say oh, we don't have a good thing and therefore we can just put them in a big group and not worry about it.

I think, first of all, there are plenty of measures of quality that are well validated and do measure what we think they measure.

And so I guess I think it's a bit confusing or distracting for our charge to say we don't have a good quality measure and therefore my -- the thickness that I draw the tiers or whatever is some way affected by that.

I can name plenty of good quality measures that work and that should be analyzed and I wouldn't be too worried about.

And in terms of the idea of cost and quality there's plenty of studies that show -- mine being one of them from a few years ago on aggressive medical care that aggressive medical care improves your outcomes. And so it costs more but you do better. So we can debate that.

But the bottom line I want to say is about your comment was that I think it's -- we are doing this under the assumption that what goes into the cost and what goes into the quality are reasonable measures. That being said, we have to decide what we're going to do.

My second point is on what Joyce stated. And I think there's some confusion. I just want to -- at least in my mind it's confused when we talk about thresholds.

I thought the purpose of thresholds in the context of the efficiency argument. And correct me if I'm wrong, but the purpose of thresholds is to say if I'm going to -- I'm worried about efficiency because I'm worried about that case where you get someone who's really efficient but giving terrible quality.

So I want to build a threshold to say I'm not ever going to -- I'm not going to give an efficiency measure to someone who's got clearly too low quality.

Now, I don't agree with the threshold. And like I said many times, we should just give all the data and not have a threshold. Because I'm worried about, well, once you make it past threshold what happens.

But the purpose of the threshold is really just to ensure that the efficiency measure isn't misused. But my worry is that it's not good enough, that the threshold

isn't a good enough tool to keep us from misusing a pure efficiency measure.

So I think that's where the thresholds came in. Maybe I'm wrong, but.

CO-CHAIR DUBOW: I think you still have to decide where it is. You know --

DR. SILBER: Or not use it because you're going to give all the data and leave it at that.

CO-CHAIR DUBOW: That's right. And then let the value of the user make a determination about what's good enough.

DR. ALMENOFF: Yes. I think I was trying to be a little more provocative than serious. But we just in the last couple of hours just keep circulating around this same.

If we're going to focus on a quality and efficiency measure we seem to need to move forward on it. Instead of everyone is coming up with every excuse not to.

So, I'm not saying I'm actually an

advocate of doing that, it's just everyone keeps coming up with side points and we're not kind of focusing on what we really need to get to.

So I think maybe that's maybe a little misunderstanding of what I was trying to get to.

But it was really to be a little more provocative to get people to start deciding how we really need to move forward as a framework. Because right now I'm very confused about where we are. Maybe many of you aren't but I'm still confused.

CO-CHAIR FLAMM: So, here we are. One of the things that we are going to do. We're going to take a few minutes and ask for public comments.

But we are going to transition into next steps and wrapping up what's going to follow after this meeting. And we have given some thought to another way to get another round of clarity as we sort of see

another pass at this in writing. So I won't give your whole punch line away, Erin.

But let me stop and ask the operator, please, if there are any public comments on the phone.

OPERATOR: Committee members, if you have any comments please press *1 on your telephone keypad. There are no comments at this time.

CO-CHAIR FLAMM: Thank you. Are there any public comments in the room? Okay. Then let me turn it over to you, Erin, to walk us through. Is that the plan?

MS. O'ROURKE: Sure. So I think we realized that there's a need for the committee to see this draft paper one more time before it goes out for public comment to make sure we're in agreement on what we're releasing out into the world.

So we're going to aim to get you a draft of the -- a revised draft of the white paper on June 2. We've asked for a quick

turnaround to get your feedback by June 9 so that Chris and Andy can quickly incorporate that to release it for a June 16 public comment.

That will be open for one month. Then we'll be looking to reconvene with you via web meeting the week of August 4 to review those public comments and decide how we want to act on them.

We'll then aim to take the draft guidance to the CSAC at their September 9 meeting to meet an October 1 deadline for this final paper.

CO-CHAIR FLAMM: The question was when will we get the draft paper. She said June 2. So about a month from now.

MS. O'ROURKE: We will be following up with you to get the August call scheduled. I think for the June feedback we'll do it via email. Unless, do we want to aim for a call?

MS. WILBON: We can try for a call

and just leave it open. If people can come, they can come. If not, they can always send their feedback via -- we'll take written or those who can participate on the call. So we'll look to see when we can do that. And don't, you know, stress if you can't attend the call. We'll welcome your feedback written as well.

So we'll see what we can pull together. But I think a discussion would be useful but I think written comments are always very helpful as well. People tend to be a little more concise in their statements and that's always really useful as well.

CO-CHAIR DUBOW: And if you could share comments if you send them by email so we can all see them.

MS. WILBON: Yes, we'll make sure that there's -- actually there's a place on SharePoint I believe that that can be shared, or we can start a distribution list.

MS. O'ROURKE: Yes, we do have the

SharePoint discussion board up so people are welcome to share thoughts there. Or if you are would prefer an email distribution list we could start that and everyone can reply to each other if that's more efficient for you.

MR. AMIN: Chris or Andy, do you have any other sort of closing thoughts? And then we can obviously turn it over to the chairs. And then we might have some here from NQF staff.

DR. RYAN: I wonder if the committee could agree to pretend that this never happened.

(Laughter)

DR. RYAN: No, I think, you know, I'm not going to bother trying to summarize everything, but I think something that we heard repeatedly was this concept of principles. And I think my feeling is that it's going to be a good way to revise this. If there's more agreement or potential

consensus around principles rather than specific, more precise guidance. And so that's going to be -- I haven't talked about it with Chris, but kind of an approach towards systematically revising this in accordance with the comments that we've gotten here.

Again, thanks for everything. We appreciate it.

DR. TOMPKINS: I'd just add my thanks to all. It's a lot to think about and we hope that we can reflect a lot of this in the next draft so it will be a useful step forward. Appreciate the meeting. It's been great.

CO-CHAIR FLAMM: Yes, just to add appreciation. I think everybody has been extremely dedicated, brought a lot of energy and passion to the discussion. Great, really great feedback and information. So you've given us all a lot to work with. We'll keep working through the summer to

take this to the finish line.

I don't know if, Joyce, you'd like to add something?

CO-CHAIR DUBOW: I just want to thank you all. It's been terrific meeting you.

I want to thank my co-chair. It's been a pleasure to work with you, Carole. And the staff and the authors. You have been enormously flexible and receptive and respectful. We don't know what you're thinking, but --

(Laughter)

CO-CHAIR DUBOW: But just -- but you know, your insights have been, everybody's insights have been so helpful. And I just want to say thank you very much for your participation.

MS. WILBON: I think from the staff perspective again I think we just want to echo the chair's sentiments and thank everyone for coming with open minds and

really being fully engaged and participating.

And I think from our perspective Taroon mentioned when we first started yesterday how we've been working with the Resource Use Steering Committee and have this idea about doing work on trying to figure out how to link cost and quality measures.

And I can say that starting this work out. And I think we've definitely made progress. And I think seeing how the conversation went and kind of looking at how the discussion kind of evolved organically and took us in directions that we hadn't really been considering. And I think that's a really good place to be for us.

I'm glad that we were able to kind of be flexible throughout the two days to kind of accommodate the direction that the panel was going. And I think that's going to help us come up with a better product at

the end of the day, or at the end of this process. And so I just want to thank everyone again for being engaged.

And I think we're in general really happy with the way things have ended up. So again, just thank you and we'll definitely be engaging you, continuing to engage you for the rest of this process to make sure we put out something that everyone is comfortable with. So, thanks.

CO-CHAIR FLAMM: Thank you.

MS. WILBON: And thank you to the chairs if I didn't say that for your leadership and getting us through. Thank you.

(Whereupon, the foregoing matter went off the record at 1:44 p.m.)

F

H

Q

T

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Linking Costs and Quality Measures

Before: NQF

Date: 05-02-14

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R Gross

Court Reporter