NATIONAL QUALITY FORUM

+ + + + +

LINKING COST AND QUALITY MEASURES

+ + + + +

EXPERT PANEL IN-PERSON MEETING

+ + + + +

THURSDAY
MAY 1, 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
ANN HAMMERSMITH, JD, General Counsel
VY LUONG, Project Analyst
ERIN O'ROURKE, Project Manager, Strategic Partnerships
KAREN PACE, PhD, Senior Director of Performance Measurement
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
JACQUELINE KREINIK, Centers for Medicare & Medicaid

TABLE OF CONTENTS

TABLE OF CONTENTS

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

(9:03 a.m.)

MR. AMIN: So, I would just like to first welcome everyone for their very thoughtful contributions up to this point. We are very excited for this meeting to get started. This has been a very important topic at NQF broadly and we have, obviously, had a lot of discussions across the spectrum about how to actually advance linking cost and quality.

I am just going to turn it over to Erin for a second here to walk us through some of the logistics for the day on the next slide and then we will go through some introductions of project staff and then walk through the agenda for the meeting.

Erin?

MS. O'ROURKE: Thanks, Taroon. Just a few housekeeping items. Restrooms are outside the main conference area, past the elevators on the right. We will have two 15-

minute breaks, one at 10:30, one at 1:45. We would ask that you try to refrain from stepping out until those times but we definitely understand if people need to take calls or do other business.

The information for the wireless is there, if you need to log on. Please mute your cell phones.

During the meeting, we will be doing our standard raise your tent card, if you would like the co-chairs to call on you. And Taroon, anything else you want me to cover?

MR. AMIN: No, that's it for now.

CO-CHAIR DUBOW: If you could also just make your tent cards, please, so that we can see them because that would be great. We can't still necessarily see them, but it is a little bit better.

MR. AMIN: Thanks, Joyce. And just maybe we can spend a minute just to talk administratively about how this meeting is

going to be run. I know that this is the first NQF meeting for some of you in the room. So, I just wanted to remind you that we welcome a broad discussion about the topics during today's meetings. So please, we would encourage you all to provide your thoughts, particularly since we have a very broad spectrum of stakeholders in various interests at the table. So, we welcome all comments.

Secondly, as tradition at NQF, the way the chairs will be leading the meeting with staff helping to facilitate to ensure that the goals of the meeting are completed, we would welcome you to just raise your placard, sort of on the side like this, in order to cue the chairs that you have a comment to make. And then we will help to make sure that the queue is sort of addressed in order.

And this is a little bit different than our typical consensus development projects in the sense that we are not

reviewing measures but we are providing, we are setting forward strategic guidance, not only to the field but also specifically to NQF, in order to set future policy related to measurement.

And so in that sense, we won't be going through a voting process but we will be going through a systematic evaluation of the draft report that Andrew, Ryan, and Chris Tompkins helped to begin for us and will further refine after the input from this committee deliberation.

And so that will be the purpose, essentially, of how we get through the structure of the meeting over the next two days.

I would welcome any questions that anyone has related to the structure but we will, obviously, go through a little bit more description of the agenda in a moment.

So, just a quick introduction to the NQF staff that you have probably

encountered through the course of the work here. I am joined by Erin O'Rourke on the far side of the table here, Vy Luong here on the corner next to Chris, Ashlie Wilbon -- I'm not sure how I missed the order already but there is Ashlie Wilbon to my right, and my name is Taroon Amin. We have been helping to support sort of this work from developing the initial proposal to the Robert Wood Johnson Foundation, which we thank for their support for this committee work and the work of the authors.

So, we will go quickly through the agenda for the day. And then before we get to the scope and objectives, maybe we go through the expert introductions. Before we get to that, I would welcome any introductory welcoming statements by Joyce or Carole, if you have any, and Helen Burstin.

CO-CHAIR DUBOW: I was just going to say, add Helen Burstin and Ann Hammersmith to the mix because they are two important

members of the NQF staff.

I welcome you. Taroon says we welcome your comments. I encourage you and Carole and I both encourage you to participate. It is really important that we hear all of your views. So, don't be shy because this paper will only improve with your views addressed. So, I hope this will be a lively and engaged conversation.

Thank you so much for coming. We wiped away the rain that we had yesterday so that you could arrive in dry weather and I look forward to the two days and to working with Carole.

DR. BURSTIN: I just want to add my welcome. I gave my seat to Ann for the very important next step of doing your introductions and disclosures, which she will lead for us.

I am Helen Burstin. I oversee our Performance Measurement work here at the NQF. I just want to add my welcome.

This has been a project long in coming. I think our affordability staff and cost and resource staff have been talking about it for about two years, tried to seek funding for it about two years ago, maybe even three. I think we are a little ahead of the curve and people thought we were strange by asking to look at this question of how to do this.

I think the world has come a little bit further along and we are really glad RWJF agreed that the time was now right to think about this issue. And we really do, I think, important work to do thinking this through. People talk a lot about how we are going to combine cost and quality and don't actually have a sense of how to operationalize it.

So, thanks to Chris and Andy for a great start on having a great foundation for this work. And again, I will be popping in and out but thanks so much for this really

important work.

MS. WILBON: Did you want to -- okay.

CO-CHAIR FLAMM: I will just add my welcome. I am very excited to be part of this project. I think we have a very diverse and rich set of perspectives here. And I think in our discussion, as Joyce was alluding, it will be so important today to get all of our thoughts out there but also to clarify what kind of guidance we are giving to this project. I think there are a lot of issues and nuances to sort of try and really hone in on what are the salient points we need to think of when we think about these different models.

MS. HAMMERSMITH: Hi, everyone. I am Ann Hammersmith. I am not Helen Burstin. I am NQF's general counsel. And, as Taroon explained, we will combine the introductions and the disclosures because it is a little bit quicker that way.

As Taroon pointed out, this isn't the typical consensus committee. So, for disclosures, it is highly unlikely you are going to have a conflict, frankly. But, we go through the process anyway, to be sure. And also to encourage you to think about what you have been up to in the last 12 months that may be relevant to the work of the committee today. So, you filled out a form a while ago where we asked you lots of questions about your professional activities.

So what we would be looking for you to do today is to disclose anything that you think is relevant to the work that the committee will do today. Just because you disclose something doesn't mean you have a conflict. Part of what we are trying to do here is to be transparent and understand where everyone is coming from.

So, please do disclose things that are relevant to the work of the committee today.

So, let's go around the table, introduce yourself, tell us who you are with and we will get rolling.

MR. BECKER: So, good morning. I am Larry Becker. I work for Xerox Corporation. I am a plan administrator for Xerox. And then I am a Board member here at NQF. I am also a Board member of PCORI. And I am working on a TAP at Yale on one of the measures. And I am relying on Helen to tell me when that is a conflict with something that we are doing.

DR. BURSTIN: It's not.

MR. BECKER: Okay. And so that is me.

MR. ROMM: My name is Iyah Romm. I am one of the policy directors at the Health Policy Commission, which is Massachusetts' new independent state agency responsible for implementation of our cost control legislation. In that seat there, I oversee a variety of activities, including a large

investment program for community hospitals, as well as our quality and system performance monitoring work.

I also have been, for a number of years, although I am no longer the lead staff on Massachusetts' development of a standard quality measure set through our process in the Commonwealth to seek to standardize much of the work that you are all doing here. So, I am pleased to be here and I don't think that I have anything else to disclose.

DR. LOWE: I'm Timothy Lowe. I work for Premier Healthcare Solutions. I am a principle research scientist. Most of my work is on infections and measure development, kind of metric measures, which I know it says in the paper we are not focusing on. But most of work tends to be looking at waste in healthcare.

And I do hold two patents and I have published but I have no financial stake in this.

DR. MAC LEAN: Good morning, everyone. I am Cathy MacLean. I lead the quality efforts at WellPoint.

With regard to conflicts, just potential things relevant to this, I guess I recently rotated off the Board of Directors of the American Joint Replacement Registry and I serve on the American College of Physicians Performance Measure Committee as well, and various other NQF things.

DR. LIKOSKY: Hi, I'm Donny Likosky. I am a cardiovascular epidemiologist at the University of Michigan. I work in the area of cardiovascular services, focusing on measurement of quality and cost.

DR. WOZNIAK: Good morning, everyone. I am Greg Wozniak. I am Director of Outcomes Analytics at the American Medical Association in the improving health outcomes area. I used to be in the PCPI area in terms of our roles around testing the measures and measure development. So, some of the

background is around the quality measurement and methodologies. I am also part of the project team working with Chris Tompkins on the Episode Grouper for Medicare.

DR. ALMENOFF: I'm Peter Almenoff. By background, I am a pulmonary ICU doctor. I am the Senior Advisor for the Office of the Secretary in Department of Veterans Affairs regarding healthcare value. And I also do the Director of Analytics, Operational Analytics and Reporting, which does a lot of the quality and financial analytic reporting for the Department of Veteran Affairs.

DR. SCHUUR: Jay Schuur. I am an emergency physician by training. I practice at Brigham and Women's Hospital in Boston, where I am the Vice Chair for Quality and Safety. I also serve as the Chair of the American College of Emergency Physicians, Quality and Performance committee. And I do health services research.

DR. SCANLON: Hi, I'm Dennis

Scanlon from Penn State University. I am a faculty member there and do research, health services research and study areas of quality performance measurement. So, I do work looking at what others are doing in this area and trying to make sense of it. Nothing else to disclose.

DR. COHEN: Good morning, I'm David Cohen. I am a cardiologist from Saint Luke's Mid-America Heart Institute in Kansas City and a health services researcher. I direct a research group there that spends a lot of time looking at issues of efficiencies and value in cardiovascular technologies in particular.

Nothing to disclose.

MS. CRAMER: My name is Mary Cramer. I am from Mass General Hospital in Boston. And the work I do really endeavors to improve efficiency throughout the operations from stem to stern at Mass General and its physician's organization.

DR. YOUNG: I'm Gary Young. I direct the Center for Health Policy and Healthcare Research at Northeastern University in Boston. Also a professor at the School of Business and College of Health Sciences at Northeastern University. I'm also affiliated with the VA's Health Services Research and Development Service at the Boston VA Medical Center.

DR. ASCH: Hi, I'm Steve Asch. I'm a professor at Stanford. I also work for the Palo Alto VA, where I am the Chief of Health Services Research and I direct a center called the Center for Innovation to Implementation Ci2i. Its focus is value measurement. I have nothing to disclose.

DR. ROUSCULP: Good morning. My name is Matt Rousculp. I am the Senior Director for Comparative Effectiveness Research and Health Policy Research at GlaxoSmithKline. I am an employee of GlaxoSmithKline and hold stock. I have

nothing else to disclose.

DR. NEEDLEMAN: I'm Jack Needleman. I am a professor at the UCLA Fielding School of Public Health and Associate Director of the UCLA Patient Safety Institute. I am also a member of the NQF Standing Committee on Resource Use Measures. And nothing else to disclose.

MR. STEPHANSKY: I'm Joe Stephansky. I am with the Michigan Health and Hospital Association. I fell into healthcare by accident. My Ph.D. is in agricultural economics, of all things but that gives me some background in some of the same kinds of techniques we are going to be looking at here.

As far as disclosures, I guess it is that I am simply becoming a grumpy old man very quickly and I think that is okay here.

DR. WONG: I'm Herb Wong. I am a Senior Economist with the Agency for Healthcare Research and Quality. Most of my work focused on the development of the HCUP

databases, the Health Care Cost and Utilization Project databases, developing them, getting them out, and conducting research with those databases and leading projects there. Other than that, I have nothing else to disclose.

DR. SILBER: I'm Jeff Silber. I am a Professor of Pediatrics and Anesthesiology and Healthcare Management at Children's Hospital of Philadelphia in Penn in Wharton.

I have been doing health services research for about 25 years. I am most interested, presently, in using matching techniques to evaluate quality.

In terms of disclosures, I have nothing to disclose. I am thinking about developing a few patents but haven't had any meetings with lawyers to see if it is a reality or not.

DR. RASK: Kimberly Rask. I am a general internist by training and associate

professor in the School of Public Health at Emory University in Atlanta and Medical Director for the Medicare QIO for the State of Georgia.

And I was going to disclose that Jeff and I were in the Ph.D. programs together but when he disclosed how long ago that was, I am going to deny it.

(Laughter.)

DR. RYAN: Hi, I'm Andy Ryan. I am an associate professor in the Department of Healthcare Policy and Research at Weill Cornell Medical College. I am on the NQF Standing Committing for Resource Use. And it has been a pleasure working with you so far and hearing everyone's comments. I am really looking forward to a great process over the next few days. And I have nothing, no financial conflicts to disclose.

DR. TOMPKINS: Hi, I'm Chris Tompkins from Brandeis University. I am a health services researcher on the faculty. I

look forward to today and tomorrow.

CO-CHAIR DUBOW: I'm Joyce DuBow. I am a member of the NQF Board as well. I represent consumers on a variety of NQF activities and at NCQA and other areas. There is always the need for consumers these days.

CO-CHAIR FLAMM: And I'm Carole Flamm. I am Executive Medical Director at the Blue Cross Blue Shield Association and lead the Center for Clinical Value, which involves our measurement activities around quality and cost of care. And I will disclose that I lead the Blue Distinction Center's Program, which is one of the programs listed in our table.

So, should we turn it back over to you, Taroon?

MS. HAMMERSMITH: I have one more thing to say and then I will go away.

Thank you for those disclosures. Is there anyone on the phone who needs to disclose? I don't think there is. Okay, nobody on the phone.

The last words I want to leave you with are with any conflict of interest process, we look to the committee members to make it work. So, as I said, it is highly unlikely that any of you are going to have an actual conflict, given what you are talking about today, but if you think you might or if you think that someone is behaving in a biased manner or if you think you might have a conflict, please do speak up. We don't want you sitting there quietly wondering if you have a conflict or thinking someone else is very biased and not saying anything about it. You are always welcome to bring it up openly in a meeting. You can go to your co-chairs, who will then go to NQF staff, or you can go directly to NQF staff.

So, any questions so far this morning about disclosures or anything anyone has disclosed?

Okay, thank you.

MS. WILBON: Thank you, everyone.

I don't think Andy and Chris actually introduced themselves as such, but they are the authors of the paper that is going to be the foundation for our discussion today. So, thank you guys for coming as well.

I did want to take some time to just over the agenda a little bit for the next two days so we have an idea of kind of what our thoughts are for how to get through this discussion.

So, this morning we will start out with kind of a brief overview of kind of why we are here, the scope of this discussion and the paper going forward and our goals. That will go into our first discussion at about 9:30, which will just be we wanted to give you guys an open forum, if you will, to kind of give some general feedback on the paper, give us some feedback on whether or not there is anything based on what we already outlined on the discussion guide, if there is other things we need to be thinking about, so that can kind

of add them and address them as we go on throughout the next two days. I think the way we have decided to structure that discussion is kind of go section-by-section in the paper so it is not too much of a free-for-all but we did want to give you guys an opportunity to kind of get out some initial thoughts and get the juices flowing before we kind of get into some of the more nuance discussions.

We will take a break and then come back to back to talk about some of the efficiency measurement approach discussions. That part of the discussion will kind of, we will have some discussion questions around composite measurement approaches and whether or not that is feasible based on what we know so far. We will have one of our senior directors, Karen Pace, come talk to us about some of the work that we have done there as a foundation as well.

We will follow that by a discussion around -- we will talk a little bit

about the goals. There is kind of a two-pronged approach in terms of some of the operational guidance that we are looking to get. The first is kind of around implications for efficiency measurement in public and private programs. So, the staff will start out by giving kind of a broad overview of the purpose of the discussion around different programs and applications and then we will go into a discussion with the panel about that.

We will break for lunch about 12:30 and then after lunch come back to really dive into some of the different models that were outlined in the paper and have Chris and Andy kind of walk us through somewhat of the results of their environmental scan and then have some discussion whether or not there is other things we should be considering in terms of models and approaches.

And then we will break out into various groups, four groups will be in this room. We are going to spread the groups out

into the four corners of the room. Our co-chairs and authors will kind of be travelers, if they will. They will be hopping from group to group. We have got everyone divided up already and we will share that with everyone probably before lunch or just before, so you have an idea of what group you are in. But we have divided folks into four groups, based on four different kind of applications, if you will.

And that will be followed by a report out of the breakout groups. So, we will designate someone in each group to be a speaker for the group, if you will. We will have flip charts so we can kind of take notes and get everyone's ideas during the breakout groups and we will come back and have a discussion around where each group landed. So, we will be asking someone from each of those groups to kind of be the speaker for the group.

We will follow that by a recap of

the day. And we will open it up for public and member comment. Oftentimes there are people on the phone or in the room who would like to provide input or have considerations for the committees. We will give them an opportunity to do that.

We will be looking to adjourn around five and then afterwards we have, for those that are interested, we have arranged for happy hour at a restaurant a couple blocks away called Mio. You are welcome to come and have a few drinks and relax and debrief for the day, if you like, and staff will be there to greet you.

I won't go too much into Day 2. That seems so far ahead right now. But does anyone have questions about how the day is going to go? A lot of our discussion will be guided by what we have laid out in the discussion guide. We put that together to make sure we have kind of a direction and some boundaries for our discussion today. But,

obviously, we are open to any suggestions, as I mentioned earlier, if you have things that you think we also need to be considering as we go through the discussion for the next day or two.

MR. AMIN: I would just note for those that are sort of new to the NQF process, this meeting is recorded and it will also be transcribed. And all of the committee's deliberations will be open to the public. Hence, the reason for the public and member comment, not only for those in the room, but the meeting is currently and will be, for the next two days, webcasted. Obviously, the time period that we go into breakout groups won't be. But just keep that in mind in terms of your comments. I'm sure everything would be fine but I'm just for full disclosure.

And in terms also of full disclosure, I just want to note for the social hour that we are going to have in the evening today, that we don't -- NQF doesn't have

funding to support that. So, we have the space located for you to join us but we are not -- you are on your own in terms of drinking.

Yes, do you have a question? Sure.

DR. ASCH: Will the happy hour be recorded?

(Laughter.)

MR. AMIN: It may not be recorded but it might be transcribed.

CO-CHAIR DUBOW: Only if you own an NBA team.

MR. AMIN: So, that's good for the agenda. Before we go into the project scope and objectives, I know we are at the 9:30 time period but I just wanted to point out a few things as we get started here on just some foundational vision of why we are here. And I welcome Ashlie's comments and Helen, I think, alluded to a number of this in the introduction.

But there is a two-fold effort that we are really trying to achieve here. For those of you who are familiar or maybe less familiar with NQF's work, one of the primary core functions of NQF is to endorse multi-stakeholder performance measures, consensus-based performance measures. And NQF started down the pathway of looking at cost of care measures in the last three to four years. And a number of you, and I would welcome comments from you as well, have been part of that journey with us over the last three to four years. And one of the things that has been challenging and has been an area of opportunity for us, in terms of endorsing cost of care measures, is that there is clearly a need in the country to try to look at the question of the growth of healthcare cost. And in particular, try to think about how we can think about per capita measures and episode of care measures that could be used for accountability applications.

However, one of the clear things that we have learned through the evaluation and endorsement of these measures through the multi-stakeholder process is that there is a concern that just looking at cost of care has the potential to undermine the quality enterprise that we have built over the last number of, at least the last decade and probably longer than that, which obviously is the foundation for the National Quality Forum, which is to advance healthcare quality for the country.

So one of the ways that we have thought about moving this forward, and again, the committee that is part of this work, many of you are here with us, again, I welcome comments on this, is that first we need some consensus-based, scientifically acceptable cost of care measures because measuring cost in itself is important to the community, but we really need a strategic approach on how to actually link the question of cost and quality

together. And the practical operationalizing of that is, obviously a challenge. And that is the path that we want to start with this committee, that we have certainly started with this committee. And we want to actually be able to provide very clear guidance for our measure endorsement process on how we start to endorse cost of care measures and what that process looks like, vis-a-vis quality measures. And so that is one very clear need of this work and it has a very clear application.

The second is that the Measures Application Partnership, which is convened by the National Quality Forum recommends measures for various programs. It has multiple objectives but one, in particular, is around making recommendations to HHS around selection of measures for various federal programs.

The other component of what we are trying to explore here is that the question of how to link cost and quality may be at the

measure level but might also be or might only be at the programmatic level, meaning that the way that you think about cost and quality linkages may be need to be done through programmatic recommendations.

And so the measures application partnership is generally interested in the question of whether there should be additional recommendations about how programs should be structured, in order to ensure that quality is not reduced through the process of looking at cost.

And so those are the two dual objectives of what we are trying to achieve with this work. We recognize that it has a very broad scope but we are trying to at least set a foundation for how we can think about this going forward.

So, this is not purely an academic exercise in the purpose of developing a white paper but it has very clear implications for how NQF thinks about both the endorsement of

measures and also the selection of measures for various different programs.

And so Ashlie and I, along with various different committee members felt that this was an extremely important are that needed to be explored and, again, we appreciate the support and leadership of the Robert Wood Johnson Foundation that was helpful in supporting the work of the committee and the authors. And again, I welcome comments from Ashlie or from any of the committee members who have some of that experience that they can share with the committee in terms of the foundational work of why we are here today.

Yes, Jack?

DR. NEEDLEMAN: Just a reflection from the Cost and Resource Use Committee work and its relation to what we are doing here today. At the last two meetings, the last two sets of measures that committee dealt with, we essentially had what is called the conditional

measures presented from CMS, a value grid crossed with a resource use measure grid. We are not asked to endorse the grid. We were simply asked to endorse the resource use measure. But clearly, the context for the resource use measure was in the grid and for payment related to both perceived value and perceived efficiency.

So, the NQF is in the process of reviewing these measures, even if it is not explicitly being asked to review them. And I think it is critically important that we have some guidance and framework in thinking about that.

MS. WILBON: I'll just add that in the discussion guide we will actually be having a much more detailed discussion about each of those kind of operational purposes around some of the work around the Measure Applications Partnership as well as the endorsement work that we do. So, the overview that Taroon gave is really just to kind of

give you an idea and a foundation for where we are going in the next two days with a lot more discussion to come around those two processes. And I think that is it.

MR. AMIN: So, if there are no other comments, we can just review the project scope and objectives very specifically and then we can jump into the conversation about the draft report. And I would welcome Ashlie and the authors to lead parts of that discussion on walking the committee through the various sections of the report.

So, the main deliverable for this committee is to look at a commission paper, which Andy and Chris have started us down to explore the current approaches to linking cost and quality measures, to measure efficiency -- And the way we are defining efficiency, again, is defined in the paper but really looking at the cost and quality signals -- and identify methodological challenges to linking costs and quality; defining key principles and best

practices for how these two signals should be -- how these two measures should be aligned; and provide operational guidance and recommendations for future evaluation, submission and evaluation of efficiency measures for endorsement, and I would add, for selection.

So, that is the objective of what we are trying to achieve over the next two days. And so are there any questions related to the scope?

Larry.

MR. BECKER: So, have we clearly defined two things: who is the customer for this output; and who are the ultimate end users of this output, so that we know how to form this?

CO-CHAIR DUBOW: -- as well. I think Jack's point about this train already having left the station is quite pertinent to our work.

MS. WILBON: So, I'll start. And

actually I meant to answer that in Jack's first comment. I lost my train of thought. So, I ended my comments early.

So, to Larry's question, the audience, I think is probably multi-fold, if you will. So, for the operational pieces, obviously around for MAP and for the CDP process, I think we would definitely like to take some of that, the output of that discussion directly to our standing committee and say look, this is where we would like to go in this direction. We actually have a meeting coming up with our standing committee at the end of June, which, in terms of timing, to Joyce's point, by the time this paper goes out for comment, around that time, we will be able to bring some of the work of this group back to our standing committee and say this is some of the guidance. How would we like to operate as a committee in terms of the endorsement work going forward. So, that is one audience, if you will.

The other audience, in terms of the measure applications partnership, obviously, would be for staff and for those committees to consider some of the guidance and whether or not the selection criteria that they have might need to incorporate some of the considerations that this panel has put forth.

And I think the other audience, potentially, would be measure developers. So, those that are out there kind of doing this measure development work and are looking to explore and expand the space around efficiency measurement, what should they be doing, like these key methodological challenges, what should they be considering as they are developing these measures and trying to put measures out and also bring measures in for endorsement.

And so, I don't know if there is probably other audiences.

DR. BURSTIN: I would probably add

one more. I think some of the early work the train has left the station. There is a whole lot more work to do. So, I think end users who are really looking to think about how they are going to combine these two and put them forward to clinicians and others, I think, will really potentially find this very useful.

I think there is going to be a lot of mid-course corrections, such as we have already seen in quality over the years. So, why would we not expect to see something very similar around cost and then the linkage of cost and quality?

MR. BECKER: I was thinking about so at the end of this, when we have great measures, who is going to use those measures? There are a whole bunch of stakeholders. And if we have that in view of how they are going to actually be able to use these and the way in which we develop them and the way in which we communicate them when we develop them.

CO-CHAIR DUBOW: You know, I think

that that need has been taken into account, in terms of how the agenda has been structured and in terms of how the breakout groups will take into account the work of Andy and Chris, because that is really what we are going to be focusing on. We are going to be thinking about use cases. So, stay tuned.

MR. AMIN: Okay. So, let's get started with really just starting with the deep dive in terms of the general feedback on the draft paper.

The basic question here is based on the stated scope of the paper, which we just talked about. I guess one of the first questions we want to start with is are there any other issues or topics that should be considered in addition to those that are indicated in the guide?

MS. O'ROURKE: Oh, Taroon, could I interrupt you before we get started?

MR. AMIN: Please.

MS. O'ROURKE: I wanted to give

Christine a chance to introduce herself and if she has any disclosures of interest.

DR. GOESCHEL: Thank you so much. I apologize for being late. The first spring in an Arlington home and I met water this morning when I went downstairs. So, I apologize.

I'm Chris Goeschel. I am the Assistant Vice President for Quality at MedStar Health. I really have nothing to disclose, have served on one previous NQF panel. I am delighted to be here.

I will say part of my intrigue is that many of our hospitals are in the State of Maryland and most of you may know that the State of Maryland does something unique in terms of cost and quality in linking yet to be defined. It continues to emerge. So, delighted to be here. I'm happy to answer any questions or concerns that folks might have.

DR. MAC LEAN: All right. So, to kick it off, nice draft. I think, though, I

want to call out the definition of quality that is in the draft. And I think that if we are going to be moving forward and having a solid definition of efficiency in measures, we need to start at first principles and have a strong foundation.

I think sometimes people get a little tangled up with the six aims that the IOM laid out. And the six aims are not the definition of quality. Healthcare quality is what we do to improve health outcomes and functional status. And if you go back to the Chasm Report, those aims were kind of the second objective. And the first objective was, in fact, to improve the health and functioning of the U.S. population. And the aims were laid out as the things we need to do to get to that desired result of improving health and function.

So, you know, as you saw in the draft, using those aims as the definition of quality gets a little circular, since one of

those aims is to improve efficiency.

So, I would like to kind of pull back for a second and think about how we define quality. And I would propose we get back to the original IOM definition. And even back in 1990, the IOM had a report prior to the Chasm and even getting back to Donabedian.

Quality is what we do in the healthcare system to improve the health or the functioning of the population. And then that way, I think the efficiency definition gets a lot more easy. And then it is cost per unit outcome and you can define outcomes in a variety of ways. But I think we need to get that part right.

CO-CHAIR FLAMM: Jeff.

DR. SILBER: I absolutely agree. It was one of the points that I wanted to make, which gets at the question of process measures. Do they have any validity in trying to state that we are looking at quality measure? We really want to look at the health

of the public and the health of the patient and the functional outcomes. And so process wouldn't be part of that numerator.

So, I absolutely agree with what you are saying.

CO-CHAIR FLAMM: Jack.

DR. NEEDLEMAN: I've got to admit, I really wanted the paper to be much richer. And there were at least three different ways in which I thought the conversation in the paper really needed to go further to enable us to do the kind of work we want to do.

One of them has to do with the definition of efficiency or value and whether we are measuring value strictly by the relative cost of producing services in different settings or whether, as Timbie and Normand do in one of the papers that is widely cited in this paper, a cost effectiveness study, which puts a value on the outcomes gets imposed on top of simply looking at the relative cost of the different providers.

So, that wasn't sort of addressed. When we talk about efficiency or value here, how broadly should we be looking at it? Should we be limiting it to the relative cost of services -- relative cost of production in the settings?

We have got other issues in the way we measure the relative cost of production but I will save that for later.

The second area where I thought the paper didn't go and really address critical issues is in the lack of consistency in ranking or classification across different approaches. When we see an efficiency measure, we have seen lots of repetition within a given approach and say do we get a consistent ranking there. But this whole set of papers, and again, some of the papers cited I am heavily influenced by the reading of the Timbie and Normand paper, really underscore the fact that you don't get consistency in rankings across different methods.

So, if that this the case, what are the implications for an organization like NQF who is reviewing measures in terms of the reliability and the need to assess whether the specific method of ranking you have picked on the value dimension influences the results. And I think we need more discussion in the paper about that implication and how to think about that.

And likewise, not enough discussion about the choice of outcomes. If they are all highly correlated, 0.98, who cares which one we choose to rank? We may care about how we do the ranking. But who cares? We are going to get the same answer. But they are not highly correlated.

And if we have got a specific efficiency measure that is built around deaths, or around infection rates, or around 13 different items, each of which has a very different ranking in itself, how we put those together and how we think about what we count,

what we don't count, how we weight them, critically important. It is mentioned in there but we need a lot more analysis about how to think about that in terms of a measure development process.

CO-CHAIR FLAMM: Steven.

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

DR. ASCH: I agree completely with the sentiment of the previous speakers in footnote that there is a circularity in using the six aims as the definition of quality.

I guess I don't agree that we should exclude process measures. This comes both speaking as a practicing clinician and, of course, as a health services researcher, because process measures are what we have under our control. And as long as we imbue those process measures with a strong link to outcomes, perhaps by including definitions of appropriateness associated with them, I think that having them in-linked quality cost

measures actually will guide the system in the right way.

So, I urge us to think broadly in that regard.

CO-CHAIR FLAMM: Gary.

DR. YOUNG: Well, my point actually speaks to this already sort of a growing debate about whether to use process measures or not.

I thought the paper did a really nice job outlining some of the different approaches. And the fact is right, payers, purchasers are already developing approaches to combining cost and quality measures.

I think another step that may need to be taken is to develop a framework in helping purchasers select them on different approaches and that would take into account some of the points that Jack just made. I think also we need to think about how well these different measures can discriminate among providers relative to the goal of

promoting health, improving populations' health. Thinking about the clarity of the information that would be communicated to providers through these different approaches, potential behavioral responses. I think we need to think about the potential motivational qualities that could emerge from each of these different approaches.

So, I think we sort of need a framework as a starting point to help purchasers select among the different approaches.

CO-CHAIR FLAMM: Greg.

DR. WOZNIAK: I would echo Steve's comment around the support to have process measures, as well as more population health level measures. Both are important but again, process measures are more controlled.

My point was more generally around how actionable are some of these measures that are combinations. It is like some of the challenges with working with composites.

Composites have value in terms of high-level population levels but they are not very actionable. You have to get to the actual pieces of the measures or the potential. The components are the measures that you have control over.

So, I think actionability of some of these approaches that are laid out in the paper might be an area that could be expanded upon. Again, it addresses the usefulness and the usability, which is one of NQF's sort of standard criteria for their measures.

CO-CHAIR DUBOW: I just want to observe that it depends on the end user. When we think about composites and the end user is a consumer, for example, the fact that something is rolled up makes it -- it simplifies the decision. So, all the more reason that the aggregation and the composite is a valid assessment of performance but it depends who is using it.

CO-CHAIR FLAMM: So, Cathy, do you

have another comment?

DR. MAC LEAN: Just to wrap up the quality discussion, I agree that process ought to be included. I would kind of expand that to even include structure in some cases, the issue being, as long as the processes and the structures are clearly linked to the outcome. That said, I think that it would be valuable for this group to make recommendations with regard to the valuation, not in monetary terms, of the quality metric. And so, I would just call out that in some areas we have lots of process measures, appropriately so. It is very difficult to measure outcome measures and lots of things may impact those outcomes; for example, within chronic diseases. But the set of measures that we may have available for a given condition, you know, they may be kind of weak, actually, on a process level and really don't mean the same thing as the outcome that is more difficult to measure.

So, I think we need to be

cognizant of the fact that if we are getting into a quantitative measure of efficiency, to understand that that quality metric has different meanings. And so as long as you comparing apples to apples, you are fine but we need to be cognizant of that.

CO-CHAIR FLAMM: Dennis?

DR. SCANLON: So, I have some similar comments. Maybe I will frame them in a different way. I think what I heard Gary say is this may be kind of crying out for a little bit of a framework that addresses the complexity of a number of issues that I thought was sort of missing from the introduction of the paper. And I am sort of torn because I wonder how much the charge of this effort is sort of directing the paper in a certain area. Specifically, around the goal is to link measures of quality and cost for the purpose of measuring efficiency. So, there is some implication that we need to take what we have, the existing measures and link

them to come up with efficiency. As opposed to, and I think I made these comments back in February, to think about sort of what we want to measure and sort of what we need to sort of get to the goal line, if you will. If we want to really start with efficiency, what would do, rather than necessarily start with what we have. And I understand that is kind of a dicey situation because we have what we have and we want to use them.

But in terms of some things I think that would be important to address in a framework, some of which have been mentioned, is this issue of are we taking what we have or are we sort of thinking about, the paper mentions briefly sort of the economic concept of efficiency but it kind of washes it aside and says we are not really going to address that.

But that sort of gets to the second point, which is, ultimately, what is our goal? And there are a variety of goals

here. If our goal is to sort of create a healthcare system that provides value and operates efficiently, then I think the concept of economic efficiency is probably an important one. And what we might be trying to sort of do through measurement is for innovation spur folks who are thinking about different ways of delivering care that helps patients.

You know somebody mentioned end users. There is a variety of sort of end users of this. I mean consumers for making choices about where to get care, particularly in a world that seems to be coming more high deductible, kind of health plan oriented. There are payers for thinking about centers of excellence like either Home Depot or Lowe's did in terms of Center of Excellence with the Cleveland Clinic and thinking about making a choice like that.

There is certainly plans, thinking about tiered networks.

So, I think outlining sort of those goals and sort of thinking about this measurement might or might not help or where some of the nuances are important.

I also thought that when you look at sort of existing consumer products, take cars or take electronics, for example, and some might disagree with this but it is not clear to me that the measures are necessarily linked. They are presented side-by-side. You have got the price and then you have got a whole bunch of information on attributes. But often, it is a third party that makes a judgment about value.

So, if you get to CNET.com or you go to Kelley's Blue Book, or whatever, you get the facts. The quality measures and the attributes and you get the price. And then usually, it is some other pundit or sort of organization that sort of makes a judgment about what the best value is in the sedan range, mid-price or something like that. So,

I think that begs a question as well is do they really need to be combined or do we just need to sort of do each other well and present them. I guess that would sort of relate to the side-by-side approach that is referenced in the paper. But that is, I think, a central focus because the task here seems to suggest that we need to combine them, as opposed to just provide the information.

And the last thing I will say is I think that the role of context, particularly in the context of healthcare markets in pricing, as it relates to cost, I think it is a critical one here as well. I go back to that Jamie Robinson paper in Health Affairs, where they look at the impact of reference pricing in California, and prices converge pretty quickly on hip and knee replacements in California. You know I guess a question is would we call the providers, who lowered their prices because of market competition inefficient to start with and now efficient

once their price went down or were they just as efficient from a production perspective. It just happened that market forces reduced prices. And I think that whole issue is an important one to grapple with as well.

CO-CHAIR FLAMM: Thank you. Before we continue with the discussion, we had two committee members join. If they could go ahead and introduce yourself and disclose any conflicts that you might have.

DR. PANTILAT: Good morning. Steve Pantilat. I am a professor of medicine at U.C. San Francisco where I direct a palliative care program. I don't know if there is more introduction or if that is sufficient. Great. Nothing to disclose.

DR. SPEIR: Good morning, I am sorry I was late.

CO-CHAIR FLAMM: Microphone, please.

DR. SPEIR: I've been sitting on the Beltway since about 8:00 a.m. So, I am

very glad to be here. I am Alan Speir. I am a practicing cardiac surgeon in Northern Virginia, representing the Society of Thoracic Surgery and I have no disclosures.

CO-CHAIR FLAMM: Thank you. Please put your cards forward. Thanks.

MR. AMIN: One other quick observation. You need to turn off your microphone after you are done. We can only have two at a time. So, that might be the reason why people are having a difficult time.

CO-CHAIR FLAMM: Go ahead, Christine.

DR. GOESCHEL: Great, thank you. I just want to say I think we are clearly on the right path with the richness of the discussion. I would just like to add a note of caution, and I think we have said it in a number of ways, the need to be really crisp and explicit in our audience, in our definitions, in our intentions.

I think it is also important, if

possible, to declare what this is not. Because what will happen when this is available is that people that we didn't intend to use it will take and use it for purposes that were not intended. And as much as possible, I think it is important going forward to keep that in mind so that the case that is articulated is kind of working with both of those models.

CO-CHAIR FLAMM: Thank you. We have Matthew next.

DR. ROUSCULP: Sorry, I was still thinking through a lot of this as comments, a lot of great insight. So, I appreciate that.

I guess for me, in reading this, instead of echoing what everyone else had already said, which I am in agreement, I applaud the authors' willingness to kind of write this and now come in front of individuals and kind of hear about our baby. It is nothing, I think, personal. I think this is a direction that hopefully we hope all

this just continues to get better. I thought this was a very good start.

I guess for me just from a process element, I notice that as far as the literature that we are going to review was primarily in the PubMed area. And with Joseph sitting to my left, I am really kind of sparked to say there are the valuation of the public goods that other areas have kind of focused on and thought about and the linking. Because, again, I don't know if it is as much.

Because earlier in this discussion with NQF, you have groups that are thinking about the cost and you have individuals thinking about the quality areas but it is more about that linkage and how do you kind of compare the two. And there is a lot of work done. And I would even echo going to the Federal Register and reading Ken Arrow's discussions that he had related to the Valdez Exxon.

I would also say then a second

area, I promised I wouldn't echo but it is really this hierarchical issues. The idea that a lot of the discussion we had was more about the organizational perspective and we kind of gloss over it and we say well, that is the same thing with individuals. And I want to make sure that, as far as the linking, that when we start talking about signal to noise type issues, that that is really going to play a lot of difference.

And we can just use an example which is, at the very front, you talk about costs and costs for patients. That is out-of-pocket costs is growing as far as an issue for procedures being done in the hospitals but for outpatient or for drugs or other things, outpatient is just a lot higher percentage or proportion and not capturing that is really is going to cause some issues.

So, I just want to make sure as far as our linking that we don't just kind of wave over and just say everything is the same,

no matter what level of the organization or how hierarchical these issues are.

CO-CHAIR FLAMM: Thank you. Coming back to Iyah.

MR. ROMM: I would echo Matthew's comments, both about the authors. I think that on one hand it is very easy to sit and provide all sorts of constructive feedback but it is extremely helpful for me to see the various models in play, which are few. And I think I am struck, in part, by the abject limitation nationwide of data to draw upon as we start this work.

I am mostly struck, coming back to some of the comments earlier, that we seem to make a leap almost immediately from linking cost and quality to efficiency. And I think we just pass over value entirely. And I think that that is a real opportunity for this group to start to think about that. Because I don't think that value and efficiency are the same thing for every audience. I think in some

settings they may well be. But especially from a consumer perspective, I think that they are considered and thought of very differently.

And I agree with the comment that as we think about linkage, not only do we need to think about different models of linkage but even with the same self-measures, the way that they are linked and presented is probably different for different audiences. I think there is a whole space there that we haven't begun to tease out.

I think that the other thing that I would just note is that I am almost struck by the fact that though we are having important and complex conversations about certain segments of care, we are just scratching the surface. And as we continue through this process, I think it is important when I look at the examples out there and I am most familiar with the self plan and the AQC in Massachusetts and others. It is entirely

hospital and a little bit of physician-based. We are not talking behavioral health. We are not talking long-term care. There is sort of a world of consumer-oriented area of focus that I think that we need to be thinking about.

So, I would encourage us to push a little bit further into that frontier as well, even if there is not much work done yet.

CO-CHAIR FLAMM: Larry.

MR. BECKER: Thank you. So, I also think that we need to build into this the thought about the fact that patients have preferences about the outcomes they want and how they can actually obtain those outcomes differ based on their circumstances. You know whether it is their education level, their socioeconomic status, and also they are going to make more and more choices as they are paying for more and more of the care that they received.

So, when we start to think about

linking these in value, to whom is the value, where is that perception? Who is perceiving the value? And I think we need to talk about that and how this all gets formed.

CO-CHAIR FLAMM: Jeff.

DR. SILBER: I can see that most don't agree with my statement about process. I would say that I haven't changed my mind and I won't. But if we do go down that road, I think, at least, and I heard some people say, that it is very important to define the guidelines for using process such that we don't crate the tautology that I am worried about between process and cost. And secondly, that we don't confuse the users, which I always think of as the patient because they won't necessarily understand why a decrement in process is worth the money that they are spending.

So, I am just adding a warning. Clearly, people want to put process in but I would like there to be some strong guidelines

about how it could be -- about not to misuse those particular aspects of the process.

CO-CHAIR FLAMM: Herbert.

DR. WONG: So, let me begin by thanking the authors of the draft. Get the straw man out there. There is a lot of concepts are floating out here really designed to help shape and to improve it.

Many of the comments that have already been stated I share a common view but there is a few that I think is worthy to emphasize, to make sure that we kind of get this draft right.

One of the things that I think strikes everyone is that this is an incredibly complex area and that things that have been stated out there in terms of different perspective, different audiences and things of that nature, I think needs to be somewhat highlighted in the draft. There is just this pure recognition and that we need to probably hone in on certain aspects of it.

So, one of the things that constantly comes to mind is the RAM report that looked at the topology of efficiency that was funded by the Agency. That document probably should be highlighted to at least lay out the concept of efficiency. So, from the economic world, clearly, there are economic terms that sits behind that efficiency definition that economists and folks that work there are very familiar with.

As the concept of efficiency emerge, at least in the healthcare environment, it is well accepted, the term health or economic efficiency but it is not really accepted in the field.

So, my caution is to recognize the complexity of that. Many people interpret it in a certain way and that that should be well-documented.

The other component I think that needs some highlighting is that oftentimes when we talk about efficiency and value, they

are somewhat intermingled. And in the end, if one were to step back and to use the analogy that Dennis presented, value almost by itself is subjective.

So, if you use the car analogy, you can have a great metric of safe quality. You have a great metric of cost. And supposed you are buying a Cadillac versus a Smart car. The measure for the Smart car could be actually higher. It is a 2.0 or whatever versus the Cadillac but they are fundamentally very different. And someone could look at this and say from a value perspective, the Smart car gives me that better value. But really, I don't want to be driving a Smart car. I really want that Cadillac.

So, again, it goes back to one of the concepts that we have all played with is that perspective matters. So, whether it is the consumer's perspective or the payer's perspective and things like that.

And sometimes a single measure

will give us a signal. And it might be in totality of other metrics that help us make whatever decisions that we need to make.

CO-CHAIR FLAMM: Okay, next I have Tim.

DR. LOWE: Those of you who are clinicians will have to forgive the engineer scientist coming at this from a very different perspective.

I think one of the problems with the modeling that I see is that we assume that cost and quality are somehow separate and that we can -- when I look at the models, I tend to see cost put as a dependent variable and that in independent variables, we add measures of quality, whether that is mortality or length of stay and we simply try to understand how it works.

I don't think that they are separate. I think cost and quality go together. You can say there is a nexus between the two. And the relationship between

them is not linear. As one goes up or one goes down, sometimes they go up or down, I think it is much more complex than that.

Going back to the electronics example, since I am an electrical engineer, I can build a circuit board for $2, especially if I have it built not in the United States. But I can built it for $2 and the focus then could be maximizing profit. Or, I could build that board much more robustly and say instead of having planned obsolescence with your refrigerator or dishwasher, I can make it last much longer.

The thing is that so there is a production cost involved in this. How much I spend on that board has to do with its quality or its longevity. The interesting thing from a production's viewpoint is that the difference between the $2 circuit board, which can burn out in a few years, and the $4 circuit board, which might last 20 years is a very small cost, in terms of production.

But the pricing, and I think we got at that, is a very different thing. So, when we talk about cost, we are not just talking about the production cost, and this is where I agree with you, Jeff, because the process measures that we have can be valued for each something that a doctor does, he or she, there is a charge for that. And that has a direct relationship to a patient's outcome. But there is a production cost and there is how we price that. And the pricing seems to be driven more by the market, which I think I have heard, too. How that is competition or how we decide to price it, which from a manufacturing view is whatever the market can stand or how I want to relate to my consumers.

And I think that is different in terms of value. Is somebody willing to pay more for a better product or is pricing more an issue and so, I will choose the lesser product because maybe that is what I can afford.

CO-CHAIR FLAMM: Cathy.

DR. MAC LEAN: Two comments. One, this is really complicated. So, I think we would be well-advised, perhaps, that the outcome of this effort is to lay out a meaningful framework. At the outset, I don't think that there is a single measure that is going to be valuable across all the different settings that we are going to be able to define. So, maybe we ought to be thinking about what is the framework for the context.

And then two other kind of related comments. This process outcome question, that is kind of the Holy Grail in quality. You shouldn't have a process measure, unless it is related to an outcome. And mathematically, you could do the math. You would have to build in a lot of assumptions to say well, I am a rheumatologist so I will go with something that I am really familiar with. So, one of the process measures is people who have rheumatoid arthritis should be on a certain

class of drugs call DMARDs. And the reason for that is because we know from a large body of literature that people who are in DMARDs have much better outcomes defined by standard outcomes for health and function and so on.

So, you could, mathematically kind of model it out to say okay, so if someone is on a DMARD, this is how that would translate into an outcome. If you wanted to go to that effort, I think there would be a lot of assumptions. And depending on the particular process measure, condition, it could be more or less assumptions. So, that is why I was saying before, not all quality measures are equal.

So, with regard to that process, outcome piece, that is something we can think about.

The other piece I just want to really emphasize is that the context is so very important, particularly when we are getting to this efficiency piece. And if you

take an example -- let's take knee replacements. So, someone references the CalPERS reference-based pricing experiment in California. Take a total knee replacement, if you are the hospital or hospital system, you have got a bunch of different surgical teams doing the procedure, you can look and see, okay, for this outcome, and we can measure function and functional outcome with that procedure, this is how much it costs for this group of surgeons to produce it, for this group of surgeons to produce it, or for this center to produce it. And so, as the hospital administrator, you can define which is the most efficient.

As the payer, you can look across and say well, how much does it cost me to purchase this as a health insurance payer. And then in that example, CalPERS, which is self-funded, said this is crazy. There is a five-fold variation in the cost of knee replacement surgery in the state. And so that

is clearly having to do with negotiations. Some hospitals are way better at negotiating a rate.

And you get to the member's side of it, if you have got a $20 copay versus a $30 copay, if your usual out of pocket is $6,500, the consumer's value proposition of the efficiency is very different to that person, until you put in place reference-based pricing and kind of upturns the apple cart.

So anyway, the context, I think, is critically important to layout in this framework that we are going to be talking about.

CO-CHAIR FLAMM: Okay. I have Donald, then we are going to Joseph, back to Jeremiah, and back to Iyah. Donald.

DR. LIKOSKY: Thank you. I guess the only thing I am going to highlight goes back to the comment about perspective. And I am a little bit confused about whose perspective we are focusing on.

For instance, I guess in the introduction piece, the purpose we speak specifically borne by the payer. But then in the definition of terms, cost speaks to payer or consumer.

So, I am trying to understand exactly when we are talking about cost whose perspective because it appears to, at least on the surface, to be unclear.

CO-CHAIR DUBOW: You know I don't know if the staff wants to address this. Generally, NQF measures look are supposed to be appropriate for quality improvement and accountability, those two groups public reporting. So that provides a range of perspectives. The accountability includes payment, public reporting, so it is a broad swath, as well as quality improvement, which means it has to be meaningful and relevant to clinicians as well.

So, that is typically. I don't know if you want to add anything to that.

MR. AMIN: Yes, Joyce. The only other thing that I would just add is that as we go through the later part of today, part of the use case really defines the audience as well. So, when we look at this question of how to link or looking at the question of how to measure efficiency, it will often partially depend on what the use case is, whether it is for quality improvement. The audience there would be providers. For pay for performance applications, that may be different, much more consumer-facing.

So, it all depends on -- well, it doesn't all depend on but it is heavily influenced by the use case. And so when we go through each of the use cases this afternoon, part of that exercise is to understand it from various different perspectives, not simply one.

DR. LIKOSKY: If I could just go back to that, though, it does say that we seek to evaluate the specific case in which cost

borne by the payer is the input of interest. So, I guess I was --

CO-CHAIR DUBOW: Well, a payer is not only a purchaser. A consumer is a payer. You know out-of-pocket costs makes you a payer. So, I mean we shouldn't be narrowed to think only purchasers when we think about payers. Government is a payer. Health plan is a payer, reimbursing.

CO-CHAIR FLAMM: And it is very hard to tease apart the member component, given differences in benefit design. I think we have to think about it holistically, understanding that there is an inter-relationship between the two.

Okay, Joseph, do you still want to make a comment?

MR. STEPHANSKY: Well, I'm still concerned about whose costs are going to count here and whether we can talk about consumer costs. But those consumers have non-out-of-pocket, non-deductible, non-copay costs in

consuming healthcare. Are we going to count them or is that something, for the purposes here, we are going to leave out and just focus on the payer? We can get really wide here, really fast, if we are not careful.

I am also concerned about we keep talking about outcome measures but about the only outcome measure that I see that is really solid is death, unfortunately. And when I look at things like readmissions, I am looking at intermediate output, rather than a final one.

So, I think we have a long way to go just on that quality side, which is, I think, what Catherine is talking about.

CO-CHAIR FLAMM: Jeremiah.

DR. SCHUUR: I'll try not to repeat too many comments. I would add that I think the perspective does need to be very explicitly articulated in this difference between whose costs is critically important and it should be articulated throughout the

paper.

And secondly, this was briefly mentioned, but I think it would be helpful for the paper to have a little discussion around externalities and unintended consequences because, in many areas of care, the price cost includes social goods. I think about my world and, for example, trauma center care where the care of all patients in trauma centers is more expensive but the existence of a trauma center does have clear evidence that it improves outcomes. And so, that should be addressed.

CO-CHAIR FLAMM: Iyah.

MR. ROMM: We can go to Larry first.

MR. BECKER: Okay, thank you. So, I wondered if, for this purpose, we could all agree that the patient is true north. So that was the comment.

MR. ROMM: Out of fear of repeating myself, I am going to come back just for a moment to the written definition of

value here because back to several comments that have been made, I think most precisely by Tim to my left, this idea of preference-weighted assessment is written into the definition of value from that patient perspective. And I think that what we are all tripping around a little bit is that that may not be something that this is the right group to wrestle with that preference-weighting. And I think that there may be a need for some acknowledgment that is for us to determine how we think about and present information. And the group that is around the table, largely, does not speak to that. But that there is a need for the preference-weighting to be sort of at the fore and for patients to be the true north.

I think the other thing that I continue to be struck by and it is referenced a little bit here in the conversation around normalized pricing and certainly to the CalPERS point, that context matters very much.

And we see in these segmented, Massachusetts most notably, that this idea of splicing apart price and utilization as we think about cost is critically important and we can't talk about them as being the same, when we have just vast price disparities in certain segments.

But also, where the entities that represent many of those price disparities are, sort of core functional units, Centers of Excellence for certain segments that are not available otherwise in markets. And so I think as we think about that context that Jay just referenced, it extends beyond just the question of certain segments of, for example, trauma care, and their social impact. But truly, the availability of specialized services in certain segments of the market.

CO-CHAIR FLAMM: Greg.

DR. WOZNIAK: Well, let me say I feel for Chris and Andy.

(Laughter.)

DR. WOZNIAK: I knew this was complicated coming in and that has been reaffirmed.

I guess one of the challenges, then, is to figure out what we can do and what we can put in a paper and recommendations to NQF that they can use, as opposed to what would be a wish list. And the comment just made around preference-weighted assessments, we don't have that. There is very little, if anything, on preference-weighted anything of any of these actors. We just don't have it.

So, if you are going to build in a framework where that is needed, that framework is not going to be very useful. You are not going to be able to operationalize that.

And it goes back to comments around efficiency from the economist's perspective, as an economist, and for the others in the room. It took us about 200 years go get there. Right? So, for us to make a jump and a leap and, obviously, with

technology and things move a little more quickly these days, it is going to take us a while to get to that kind of a framework where there is a lot of assumptions around efficiency. There is a lot of givens around efficiency. It is a very standard framework that took, again, 200 years to develop. I won't go into Marshallian and all those kinds of concepts, in terms of economics. But we need to say what can we do here? What can we construct that we already have information and what is doable. And then maybe have something like we should have or, in addition, it would be nice to have and we should look to build some of these other tools, some of these other variables, into the framework and into some of these concepts.

But to operationalize some of these things, are almost impossible. Go back to value is some cost over quality or quality over -- some cost quality metric, where you don't have a measure of either of those. So,

what does value really mean then, if you can't measure either the numerator or the denominator.

So, I think we need to have sort of a reality check. What can we do? What can we operationalize? What can NQF use? What can those recommendations say that somebody developing measures would be able to apply?

And also, the comment, I thought it was very useful, Dennis' comment, the title of this group "Linking". It assumes that you have got cost measure. You have got quality measure. You are going to somehow put them together. If you ever tried to do that, it is like building composites. There is all kinds of constraints that have to be met in terms of eligibility and the populations have to match and all kinds of things have to match up time-wise, population-wise, condition wise.

I don't know if anybody has -- has anybody gone out to start developing efficiency measures where they develop both

cost and quality at the same time? I don't know if anybody has done that. And maybe Chris and Andy, in your literature review, have looked to see that but I don't think anybody has done that. I think it has always been a linkage approach. And doing them simultaneously, or at the same time, however they are done, would be a useful maybe recommendation or something to look at as well.

CO-CHAIR FLAMM: Jeff.

DR. SILBER: There was an issue that I brought up at the phone conference a few months ago which didn't get into the report but I had, and I don't know if I made the point clear enough about direct and indirect standardization, but I just want to go back to it. It has to do with perspective.

But if I looked at some entity, say a hospital and I used indirect standardization to evaluate their quality and cost and, in the end, some kind of value

metric, I would be basing my metric on the patients that that hospital saw, which could be different and non-overlapping from other hospitals that see very different patients.

The indirect standardization approach, which is generally used, has that problem that you might have non-overlapping patients. So, for an individual patient, you might see excellent value at this hospital and poor value at that hospital. But in reality, the first hospital didn't really overlap with themselves.

So, another approach is to use direct standardization or do what we can to make the basket of goods that we are comparing similar, not just through the overall adjustments, there are different ways to do this. But we really need to make sure that, as a consumer, that number one, the basket of goods they are looking at is relative to them; and number two, that the basket of goods is similar across the entities that they are

comparing.

I don't think the report dealt with that, unless I missed it. But I think it is a crucial concept that needs to be incorporated into the report.

CO-CHAIR FLAMM: So, I just want to summarize who I have on the list. And we are coming up towards the end of our hours, just to capture -- Larry, did you have another comment?

MR. BECKER: No.

CO-CHAIR FLAMM: No, you put yours down. And then I have Alan and Steven. And then Iyah, do you have another comment?

MR. ROMM: No, I'm all set. Sorry. Thank you.

CO-CHAIR FLAMM: Okay. Thank you. And I have Jack and Matthew. Alan.

DR. SPEIR: Thank you. I would like to applaud the efforts of our authors to make some sense of this. You have given us at least a structure to shoot at and thank you

for that opportunity.

I am attempting to resist the change to stay on task with linking quality to cost. And as I read through the document and then in listening to the comments, there is the seeming vacillating of using value that is thrown in in an ill-defined attempt because, as we have all said repeatedly, values being differentially defined by the payers, whether it is CMS or Medicaid, the private payers, and particularly the patient, it was pointed out before, patient satisfaction will ultimately, if we are led to believe and believe what we read up to 30 percent of reimbursement.

So, there is a disconnect to me in that the more dissatisfied a patient is, that may actually save cost from the payers by lowering reimbursement. The more satisfied they are, there may be more payment given out to the providers.

Be that as it may, there is also a disconnect, and I had hoped there would be

more focus given about the variability in pricing and reimbursement. Because, again, there may be, for one example, up to 75 percent difference in how a ventricular assist device or transplant is reimbursed on the East Coast and in the Midwest. I think if we are moving from paying for value, rather than volume, this differential is unacceptable and we, I think, as a body, need to give some deference with that.

I would respectfully disagree with the previous comment that death is the only outcome. I think depending on the specialty with your practicing, there is very real definitions and focus given to the types of outcomes by which were measured and we are actually, as institutions and providers, given star rating systems. And again, there should be more motivation given, regardless of the specialty to helping to define what we can get our hands around as outcomes.

And lastly, the last section,

there were implications for the NQF endorsement. I found that somewhat disquieting, frankly, because we are hardly coming to grips with some of the grievances around these term definitions, much less allowing a formal body to then value what attempts are being made by well-meaning organizations or societies and should that validation be delayed until we are a little bit better in consensus as to what we are about.

Thank you.

CO-CHAIR FLAMM: Steven.

DR. PANTILAT: So, from my world, as a palliative care physician, I have to agree that death is not the only important outcome, since that is a very common outcome. And we --

(Laughter.)

DR. PANTILAT: I know. And we think we provide a lot of quality in that setting. So, fair enough.

But looking from that perspective, our field has really been built on looking at issues of cost and quality. And really in thinking about the quality provided in palliative care, you can't make a buck doing it. And so the way that palliative care has grown in this country is largely on cost savings and so linking quality improvement and cost savings has been essential to the growth of this field and has been done for a long time. It is not done at the micro level for one patient but it is done really at the macro level.

And I think the comment about the trauma center, I think really applies in palliative care as well. There is actually very good data to demonstrate that the presence of palliative care services do improve quality and reduce costs. And in some ways that is a structural measure that says you can link these two in a particular way that deals with both issues. We tend to think

of it as value more than efficiency. But there may be efficiency there and there may be efficiencies within structures of palliative care services, for example, of how to make even greater efficiency, greater value.

And then the other point about the perspective, I think, is really important here, as far as the services. And I think the point about what services are included is very important. We have patients at the end of life for whom staying in the hospital cost a family nothing. That is not the quality that they want. The quality they want is actually to be home. But what they need is they need a caregiver to be there to take care of their family member, which could cost a fraction of a day in a hospital or a skilled nursing facility but there is no one who pays for it.

And so if you look at the perspective of efficiency, it really changes. But if you look globally at the issue of cost to the system, you would much rather have

someone at home. And for lack of a $15 an hour caregiver, somebody spends $1200 a day in a hospital. And so, it would be easy to manipulate sort of where the money is being spent. And I would encourage us to think more globally about it because then it allows us to really think about what kind of services are being provided. And if we take a global view, we might in fact say that there is greater efficiency by providing caregivers at home, than there would be in keeping people at the hospital.

CO-CHAIR FLAMM: Jack, thank you.

DR. NEEDLEMAN: It has been an interesting hour and a fascinating conversation. As I reflect on it against the committee's charge, I have got share Greg's reaction that I don't envy Andy and Tim.

But I think it is useful to reflect on the three different levels at which comments have been made and the implications for where the committee's work has to do. And

Andy and Tim as the agents of the committee in structuring our report, one of them is we continue to go back to the broad conceptual, fundamental issues of what are we talking about. What do we mean by efficiency? What do we mean by value? What counts? What doesn't count? And it is clear from the conversation that we have had in the last hour, we have not resolve those, nor does the paper fully provide the framework for at least reflecting on the differences we have and how to think about that. And that is one of the directions we have to go in and it is going to keep shaping the day.

But if we just stay at the broad philosophical issues, we have got two other issues that we are never going to get to that are equally important. One of them is there are a series of fundamentally alternative different approaches to linking cost and quality. Greg said has anybody tried to do this, besides the simple univariate grid. And

the answer is yes. And we have got some examples in the paper and we have got more examples of other potential ways to do this.

And in terms of those broad approaches, we really need to think about what their relative performance is. Because at some point, somebody is going to bring into NQF -- well people are already using some of these approaches to actually make payment decisions. At some point, somebody is going to bring it into NQF to say would you endorse this. And we need to understand the differences in the performance of the different approaches, that comparative across them. And we need to think about how to do that. And that is another area, where, as I said, initially, I don't think the paper has gone far enough.

And the third is a whole series of specific technical issues. Do we use shrinkage? How strong a shrinkage variable do we use? Standardized products and

standardized costing, Jeff has raised a whole bunch of technical issues.

A whole series of technical issues in terms of evaluating these measures as well that we at least need to get laid out, even if we don't fully resolve the best approach in the context of this paper.

But those three areas, the broad conceptual, the choice among alternative, fundamental models of how to do the linkage and the specific technical issues in implementing those models all need to be part of this paper, all need to be part of the committee's discussion.

CO-CHAIR FLAMM: Thank you. Matt, you are the last comment.

DR. ROUSCULP: Hard to kind of follow up on Jack.

(Laughter.)

DR. ROUSCULP: So just to build upon what Jack had just said, because I think he really hit it. When Alan was kind of

reminiscing or talking about the potential for gaming that we will see, any of us who have kids, pets, or distant relatives, we know gaming exists from a personal level.

What I really thought of value of this paper was the actual grid you had put out to say look, this whole notion of linking of cost and quality is already occurring. It is not as if we are trying to create something new. It is not as if we are trying to kind of be at the front edge of this area, it is occurring already within the healthcare system.

And there is, I am sure, an immense amount of learning to be able to go in there to be able to say look, what are these measures. What are the lessons learned? Where are the directions going? What is the good and the bad of it.

And I think in large part there is a whole lot of information perhaps that you are aware of but just because of the direction

or the length of the paper you just didn't have the ability to kind of go into all that detail.

I think there is a whole lot of value there and I would hope that you would really take that portion of the paper and really expand upon it. Because at least to me, I found that to be the most important area.

CO-CHAIR FLAMM: Great. Ashlie, if you want to close this up.

MS. WILBON: Yes, I just want to thank everyone for their comments. I think we have been feverishly taking notes and kind of having side conversations. I think there is definitely some things that we can help clarify in terms of some of the discussions around scope and where we are going and focus. We will try to do that when we come back after a break. And I do also think there are some things that come up that we are going to try to make some adjustments or some additions to

the discussion guide to make sure that we integrate some of their comments into the discussion as the two days go on.

So again, thank you for a really rich discussion. And we will break for --

CO-CHAIR DUBOW: Before we break.

MS. WILBON: Yes, sure. Sorry.

CO-CHAIR DUBOW: Well, I just wonder whether Andy or Chris has any --

MS. WILBON: Yes, thank you.

CO-CHAIR DUBOW: -- quick reaction to the assault.

MR. RECHARDT: No comment.

(Laughter.)

DR. RYAN: I just wanted to -- you know these comments were really outstanding and it is really an esteemed committee and it is truly an honor to be working with you guys.

And I think you know we apologize for the existing shortcomings of the paper. Obviously, we are not inside all your minds and we can't put everything -- you know

everyone has their own perspective and we can't address everything. And this is a really good opportunity for us to really find out what are these key things that are out of here and should be in.

And some of the points that I think we will come back to over the course of the day, this idea of the context mattering and the use case is really critical and I think our discussions with NQF over the last week about how to organize this meeting will really get at that and the relative pros and cons of these different approaches for different uses is one.

I think another issue that has surfaced, actually, in a conversation that Chris and I had over the last few days and come up today is this, to me, this idea of -- and Joyce and I were talking about this this morning is how much should an efficiency measure be specified? To what extent should it be specified before you are making a value

judgment? So, we don't want to just do a data dump of cost and quality. We want to make sense of it in some way. But what is the right -- how far should we go towards classifying efficiency in weighting measures before we are getting to this issue of value.

And we don't want to because, according to the NQF definitions, value really depends on the eye of the beholder. And we don't want say this is value. We want to have that next step be judged by the user. But at the same time, we want to have the measures be meaningful. So, I think that is an important intention that needs to get worked out.

Then one other point I would like to make and I would love to hear Chris' comments is we don't have -- there has really been basically no attempt to develop a single measure. You put cost and quality together and say this is efficiency and this is an efficiency score. I think it would be important to say why haven't we done that and

is this completely doomed? Because really it was striking to us that really no one did it. And so, I think just kind of coming up with a list of reasons maybe good or not so good as to why this hasn't happened would be, I think, a nice contribution of the group.

DR. TOMPKINS: Well, I'm going to be brief standing between us and the time-limited break.

At the very beginning when I introduced myself I said I was looking forward to the next couple of days. But the reason why we are having a two-day meeting and not a two-hour meeting is because we are trying to work together on all of this.

And the paper, as such as it is, seems to have been useful to stimulate the conversation, which is to say how far we have gotten and where else do we need to go. And so therefore this meeting provides for all of us a nice time for a mid-course correction, regrouping around what we are really trying to

intend to do and what is the most meaningful statement that we can make to the field. And as we have breakout sessions and further discussions in the next day or so, I hope that we can all come together and give us another direction.

One of the things -- now I am going to regret not just stopping there.

(Laughter.)

DR. TOMPKINS: You know I see this endeavor sort of in three categories. One is to have cost and quality measures just sort of out there. And then most obvious easiest thing to do is side-by-side. You say you want to know cost, it is over here. You want to know quality, it is over there. If you want to put them together and it makes sense, go ahead. That is sort of the costs as cost.

And then at the other end of the extreme is the attempt to actually come up with a value measure but that is actually implicit in the minds of the person who is

doing the side-by-side in the first place. So, when you bake it into and call it a value measure, what you are doing is baking in that subjectivity. Now, like it or not, that is exactly what everybody does. So, the question is, is there a scientific rigor around that or is it really just a subjective process which is unbounded and is up to the discretion of the user?

So, part of what the workers are trying to do is to decide to what extent is there a scientific framework was a word that was thrown around a lot, technical way to do it is another way to think about it.

Now, in the middle is the orphan, as far as I can tell, this efficiency measure. I'm not sure if there really is a demand for it, although everybody talks about it. Because one thing is this downside of it talking about efficiency is we bring in economists who have 200 years working on this and they say -- 200 years and counting -- and

it means a lot in the technical discipline of economics. And almost by bringing that word in, we are bringing in a field of discipline that talks about production efficiency, technical efficiency, allocated efficiency and so forth. And somebody mentioned the RAND framework and that word has a lot of technical deep-rooted meaning. And I am not sure if there is a demand in this room or in the field to bring all that in, aside from the production part of it.

But as soon as you know -- and this is my last thing -- we tried to stayed tethered sometimes. Now one of the things that tethered us, for better or for worse, in the first draft was to say that we wanted to build this out on some already established beliefs or definitions, hence, the hand wave and say this isn't about the economics, the technical efficiency and production. And rather to say the field has announced, in part, including NQF, that they have some

definition already existing about efficiency and another, somewhat distinct definition already existing about value, and what we are going to do is borrow from that and build on it.

Now, I guess so as we go through the deliberations we say, okay, is that too much, being too tethered? Meaning that those definitions that are sort of widely understood and believed, are those still useful to build on? If that is the case, then that sets aside some of the production efficiency and so on that comes from the field of economics.

Now, almost a break.

CO-CHAIR FLAMM: Those were fabulous comments. Thank you.

In the interest of our break in our schedule, we are going to start at five minutes of. So, we will take an 11-minute break. Okay? Thank you.

(Whereupon, the foregoing meeting went off the record at 10:41 a.m.

and went back on the record at 10:55 a.m.)

CO-CHAIR FLAMM: All right, I think our next agenda item, we should go ahead and get started because we still have a lot to accomplish before lunch. We will be, in the next section, focusing on efficiency measurement approach considerations. And we will have Karen Pace from the NQF actually walk us through the efficiency measurement approaches. So, Karen, let me turn it over to you.

MS. WILBON: Karen, do you mind if I just do a really quick intro? Karen Pace is our lead methodologist here at NQF and she has really been the leader in our work around many of the challenging methodological issues that come forth, one of them being composite measures. And actually Andy's parting comment in his remarks at the end of our last session is a great transition into this session, which we are really trying to get an idea from this

group on whether or not a composite measure or single measure of efficiency, in terms of combining cost and quality measures into one measure with a single score is actually feasible. Is it something we should be striving for, recommending?

And so, in order to do that we wanted to give you guys a little bit of background on some of the work we have already done in composites on the quality side. We have brought together, convened groups of methodologists in the past to really think through this in terms of what the methodological issues are and considerations and, in particular, how NQF evaluates them and defines them.

So, that is the purpose of this discussion. And so we are going to allow Karen Pace to --

MR. AMIN: Ashlie, can you also -- one other thing that we wanted to point out as part of this discussion is to consider how the

models that have been proposed in this draft report, how they can be considered, whether they could be considered composite measure approaches or how they are different. How they are similar or different than a composite measure. So, that is another area in addition to what Ashlie described as one of the things that we really want to get cleared before we move forward.

MS. WILBON: And I would just add some of that -- after this session, we are going to have Andy and Ryan do much more, a deep dive on the review of each of the models that they propose and kind of talk in a lot more detail about the environmental scan. We will break out into groups and some of that discussion around whether or not these approaches could be considered composites may actually come out in some of that discussion as well.

So, we have only got about 30 minutes for this discussion, so we will see

where we get but I just wanted to put that on your radar. So, there will be other opportunities to have this discussion around the models and other parts of the agenda for today.

DR. PACE: Good morning, everyone and thank you. I am going to basically go through the conceptual thinking of our Composite Measure Committee and where we landed in terms of defining composite performance measures and how we are looking at them for purposes of NQF endorsement. So, let's go ahead. Next slide. All right, next slide.

So, you have already seen this in terms of our definition of a composite performance measure is basically pretty simple is that it is a combination of two or more component measures, each of which individually reflects quality of care into a single performance measure with a single score. So, I guess one of the things you will have to

kind of think about in your mind, as I go through this, that efficiency is a domain of quality, at least for a lot of our thinking. So, quality here is a broad concept of quality but we have primarily seen this in the typical process in outcome type of measures.

So, one of the things to just keep in mind that we have done some work earlier in 2008 around composite performance measures but we really explicitly keep talking about a composite performance measure because, as many of you know, the term composite can be used for describing an instrument this multi-item instrument and we do not -- NQF does not endorse those and we don't consider that a performance measure in itself. So, we can have more discussions about that if you have questions but I just wanted to make that note.

The other thing, like many of these areas, different disciplines may use different terminology. So, some people refer to composites as indexes. Some refer to them

as scales. And I think in my quick read of the paper, I guess composites would be most aligned with the unconditional model. But we will go through this in terms of our thinking and you can see where you think it falls.

Okay, next slide. So, just in terms of thinking through this, the committee really spent some time on really thinking about what is a composite performance measure and really focused a lot on what is the quality construct and rationale, that you really have to have a good idea of what it is, how you are defining quality in order to really construct a composite performance measure.

So you have to have that overall quality construct. What are the components that form the composite? How are the different components going to be aggregated and weighted? And then, ultimately, the composite performance score. You come up with one score. Next slide.

So again, the committee looked at common approaches for constructing composites. And this is where you really get into different languages in terms of reflective approach, formative approach, patients receiving all necessary care, individual patients not experiencing, for example, adverse events or complications.

And I am going to go through some examples of these but one of the things, again, that the committee decided and finally came out in their recommendations for our criterion guidance is not to get hung up on these terminology but to really very clearly explain what the quality construct is and what is included, versus trying to use terminology like reflective and formative. Next slide.

So, basically, if you get into the literature about composite construction, a reflective composite is where the quality construct is seen as causing are reflected in the component measures. Sometimes these are

referred to as psychometric composites, scales, homogenous scales, dimensional. So, you have an idea of what quality is and it is actually reflected in the individual component measures. Next slide.

Versus formative is where the quality construct is seen as being caused or defined by the component measure scores. And some of the literature refers to these as clinimetrics indexes, heterogeneous index, categorical. But again, this is why we didn't want to get too hung up on the terminology because this sometimes becomes difficult to think about. Next slide.

So, basically, again, in terms of the criteria we want people to have a good concept of what their construct is and how the components that are being suggested as being included relate to that overall construct. But we really have defined, and I know you have seen these, that a composite can, first of all, be a combination of two or more

individual performance measures. And I am going to go through a couple of examples of some of these just to make it more clear.

So, for example, we have a CABG Composite Score This is a measure from the STS. And it is a Comprehensive Assessment of Adult Cardiac Surgery Quality of Care. And it actually has four domains made up from 11 individual NQF-endorsed cardiac surgery metrics.

So, I am not going to go through all of these but the four domains, there is operative care, then there is perioperative medical care. Next slide. It includes the risk-adjusted post-op mortality and risk-adjusted operative mortality for CABG and then risk-adjusted post-op morbidity. And then there is a complex aggregation and weighting methodology.

But the point here is that this composite you end up with one score and it includes the four domains or dimensions and

each of those domains or dimensions may have one or more individual performance measures.

Next slide. This is another one -- yes?

DR. SILBER: Go back to the other one.

MS. WILBON: Use your microphone, please.

DR. SILBER: Was there a rationale for why the composite score should be made up of those elements weighted by the standard deviations?

DR. PACE: Yes. In the measure submission, they do provide that justification and the rationale. I won't get into that right now. I am just trying to kind of describe what we would see as a composite. But that is exactly what needs to be reviewed when they come into NQF. In addition to our usual criteria is that we want to look at this idea of what is the quality construct and what is included in the rational for that. And

then, under scientific acceptability, there is criterion for composites where you actually do empirical analysis to justify what was included in the composite.

Next slide. Yes?

DR. MAC LEAN: Just, Jeff, I know you were going with this on this concept of can you link the process to the outcome? And we have somebody from the STS here and I can't recall but was this composite linked to an outcome?

DR. SPEIR: Well, all of those are outcomes. I'm sorry, I don't totally understand what you are getting at. But it does link. And I think what I was hearing, his question, is that it takes the total process of care from the preoperative treatment to the intraoperative care, the post-operative perioperative morbidity and then the discharge medications and the HCAHPS scores.

DR. PACE: It actually combines

process and outcome measures.

DR. SPEIR: Yes, so it is all the above.

DR. PACE: Right.

DR. SPEIR: But whether it is applicable to other specialties, it would be no.

CO-CHAIR FLAMM: So, I think we could have a long conversation as an aside. But let's bring us back to, if you don't mind --

DR. PACE: Right. So, these are just to give you an idea of different composites and what NQF is viewing as a composite performance measures. This is another one that combines two or more individual measures. And this is mortality for selected conditions. It is an AHRQ measure and it has several individual condition-specific mortality measures that they combine into getting one score from the combination of these individual performance

measures.

Okay, next slide. So, the other kinds of composite measures that we tend to see are composites that most people refer to as all or none, meaning that they are conceptualized as all the necessary care or components of care for a particular topic.

So, all the components must be achieved, meaning that you look at these patient-by-patient and if the patient is missing one thing that they should have received, then they are not accredited in terms of receiving all the necessary care.

There are other scoring methodologies where you get partial credit. The more things that an individual patient receives, of course, the more credit. So, again, there are different ways of scoring these, even when you have the concept that the composite includes all of the items of necessary care.

Next slide. So, an example of an

all-or-none composite is the optimal diabetes care from Minnesota Community Measurement. And this includes -- this is really described as the percentage of adult diabetes patients who have optimally managed modifiable risk factors.

So, these are really, this is really a combination of outcomes but it is looked at the individual patient level first. So, you are not measuring each of these individually in the provider's patient case mix. You are looking at each patient to see was the A1c less than eight, was the LDL less than 100, blood pressure less than 140/90, are they a tobacco non-user, and are they taking daily aspirin. And again, it is looked at each patient. If you are looking at the scoring methodology, if the patient has met all of these, they get a one. If they anything is missing, it is scored as a zero. Next slide.

Okay, and then kind of on the

reverse side of that, often referred to as any-or-none, these are composites that reflect non-experience.

So, usually this is about experiencing healthcare-acquired adverse events or complications. Or it could be a measure of unnecessary or inappropriate care. Typically we see these as composite complication measures. Next slide.

So, an example of an any-or-none composite is complications within 30 days following cataract surgery. And you can see here the components are listed here, retain nuclear fragments, infection, dislocated or wrong power, retinal detachment, wound dehiscence.

So, basically, again, this is looked at at a patient level. If the patient experienced none of these, that is the requirement to meet the measure. If they experience one of these, then it would trigger the scoring was different. So, it is either

any of them, any complication at all or none is basically the scoring for this type of measure. Next slide.

So, I think I will just talk about a few things in terms of the committee's thinking about these measures and how we would look at these in terms of evaluation. Again, this is in the context that NQF endorses performance measures that are intended for use both in performance improvement and accountability applications.

And the construction and evaluation of composite performance measures should be based on sound measurement science and not necessarily constrained to adhere to a specific method or categorization, such as psychometric and clinimetric approaches.

The primary concern for endorsement is whether the composite performance measure is based on sound science, produces a reliable signal, and is a valid reflection of quality. So, then that is

consistent with our criteria for any measure. Next slide.

And again, coming back to a coherent quality construct and rationale, that constructing a composite really relies on that. The TEP really didn't subscribe to the notion of there is a bunch of measures out there and I am just going to throw them all in the mix and come up with a computed score. They really wanted them -- it could be a very broad quality construct but there had to be some thinking and rationale, not just because there is a measure we can throw it together and come up with a score.

Okay, next slide. I will just -- and I think the other thing that you saw is that we did delineate what types of measures will be and will not be considered composite performance measures for the purpose of NQF submission evaluation and endorsement. We need to keep in mind that this is a pragmatic listing and we realize that there are

different categorizations and terminology out in the literature and in different disciplines but we needed to come up with a list because we have certain additional criteria that these are going to be judged by and people need to know in advance what to do, what to submit to us, so that the committee can evaluate them. And this is something we will need to continue to reassess.

Next slide. So basically, we said that these will be considered composites and these relate to the examples I just showed you. Any measure with two or more individual performance measure scores combined into one score and then these all-or-none, or any-or-none measures, that we have classified these as composites and we expect them to be submitted with the information to evaluate the additional criteria.

Next slide. And these will not be considered composite performance measures for purposes of NQF. I don't know that we need to

get into these but these are things that we do see performance measures that come in. A single performance measure, even if the data are patient scores from a composite instrument. So, this is where we have had some confusion in the past and this group really clarified it.

So, for example, CAHPS surveys. You may have a performance measure about physician communication and it is based on a multi-item instrument and the whole CAHPS instrument has lots of items, different domains. But the performance measures are really one kind of domain or concept and it is just an individual performance measure. It, by itself, is not a composite.

Now if CMS and AHRQ ever decided to put all those domains into a composite, then that would be a different story. But that was one of the things we wanted to try to clear up.

And also, we have some of our

outcome measures where the risk adjustment methodology has what some people have called reliability adjustment, Bayes shrinkage estimation, et cetera, where you are combining information from the provider with information from the average. Some people have termed those composites and the committee said no, for purposes of NQF, those are not going to be considered composite performance measures.

Obviously, we accept those measures and we have a whole lot of criteria to look at, those measures and the risk-adjustment methodology.

But anyway, I will stop there and we can see if you have any questions that I can help clarify.

CO-CHAIR FLAMM: So, we have about ten minutes for committee discussion. And, in particular, we would like to focus discussion around the question in the discussion guide, which is: Is a single score composite measure of efficiency feasible?

Okay, so that is kind of one of the things if you can touch on that. I have Larry.

MR. BECKER: Thank you. So, it appeared to me that slides 27 and 30 were similar. So, if you go to 27, I think it was, there were two conditions and you had a series of -- go to 30. No, it was the one with the retina. No, the cataract and the -- right.

DR. PACE: Keep going.

MR. BECKER: Keep going.

MR. BECKER: No, no, go forward.

DR. PACE: No, you have go to forward. Sorry.

That was it.

MR. BECKER: So, that is something I know about because over the last 12 months, I have had two detachments. I have had oil placed in my eye. I have had a vitrectomy. I have had a cataract surgery and a YAG.

And what is really important to me is can I see. And I don't see that there.

DR. PACE: You know, and again, these are very abbreviated information but I assume something about maybe the wrong power. I'm not sure. But you are --

DR. BURSTIN: It is a separate measure.

DR. PACE: It is. Right, that is a separate measure.

MR. BECKER: But you are talking about putting together a composite measure.

DR. PACE: Right, but this is a composite about complications.

MR. BECKER: Okay.

DR. PACE: And that gets, again, what is the quality construct that is being viewed for this composite? So, they started with the idea they wanted a composite of complications. And as Helen said, there is already a performance measure, which most people are most interested in the vision that they receive.

And I think that also points out

kind of one of the down sides of composites. So, say you put that in here. The question is, if that is what is most important to you, wouldn't you like to see that separately, rather than as part within a composite.

But again, these are the things that are debatable and why you have to really start out with a strong quality construct and rationale of what you are trying to accomplish with the composite and think those things through. Good points.

CO-CHAIR FLAMM: Steven.

DR. ASCH: I have a question and a comment, if that is okay. So, the question is: Are any of the other composite measures that you came across ratios rather than essentially sums, which is what these are or conditions?

DR. PACE: I don't think so.

DR. ASCH: Other than the efficiency measures, which we are going to consider today. Because I think the math

matters there.

So, then my comment, which is unrelated, has to do with the thing you told us not to talk about, which is reflective informative measures. Because I really believe that it seems like a technical point but it is a really important point because it has enormous implications for the way people actually use the measures.

Without going into my limited understanding of the distinction between the two, either you view them as you are measuring -- everything that is in the measure is in the composite or it is supposed to reflect something broader. It is an indicator of something beyond what it is actually measuring. And that indicator of something that is a broader concept is the way I believe most people interpret quality indicators. And if we don't take that into account in efficiency measures, then I think we are going to end up constructing measures that are

either too narrow or are pointing the system in the wrong direction.

DR. PACE: And I guess what ended up happening with the committee, because they kept talking about not worrying about the technical terms, but when it comes down to how you even do your analysis, it depends on how you view that, those components, whether they are reflective of or the reflective or formative model. Because one is based a lot on correlation analyses, which may not necessarily translate to the other.

So, again, you have to have a good idea of what that construct is and how those components relate to it.

CO-CHAIR FLAMM: Timothy.

DR. LOWE: I think what you are trying to do is great. And I should have started out before saying you guys did a great job. Thank you. I'm sorry I didn't say that before.

I see quality as a multi-national

construct. So, the question is can you create an index score, one number that will give you some meaning, in terms of we want to compare people. It raises some problems, though, and that is, if you want to have different subscales under that, each of these have their own demands unto themselves and each can be measured in a different way.

So, if we want to do this, I think it is feasible but you have to find a way that all of those demands can be measured on a similar scale. Now, I know people have used the -- I think you used the coefficient of variation. I saw the mean over the group standard deviation. So, I mean that is one way of solving that. But I do think people want to know because it is, going back to people's value, if I am having eye surgery, I am definitely going to want to know, especially if I am having cataract surgery, but if I am having my gall bladder, I am not particularly interested in -- I wouldn't see

an ophthalmologist anyway for that, hopefully not. Unless, I guess if it was really emergent, something on a highway maybe. But people's interest in this.

But I do think there have to be, if we are going to create one scale, maybe we can't do it. And I think you were saying probably not. I won't speak for you. But if we are going to have one number, then it is going to be very complex. Now, just coming at it from a statistical standpoint, it is very difficult to do this and I am not telling you anything you don't know. It is challenging but if you want to do that, which demands do we pick? And then how do we measure it in a way which we can combine? And if we can do that, then the answer to the question is yes. If we can't do that, then the answer is we have to find another way to display that information.

So, I am trying to answer your question the best I can but it is not

particularly an easy question to answer.

CO-CHAIR FLAMM: Okay, we have a number of people who want to speak. So, I am going to go to Peter.

DR. ALMENOFF: So, just from an operational perspective, we actually do build an efficiency model for 150 hospitals within our system and we actually have stars, which really make people extremely happy. I am saying that sarcastically. And it is a single score. But what we then do is we build sub-models. So, to just build a single score and have no action to it, is very problematic. The idea is if you are going to build a single composite, you really have to have sub-models. So, we have 14 sub-models to allow the different sites to strategically target what they are interested in focusing on.

So, if within our big efficiency model pharmacy, buying care outside, lots of different things that we look at, they can basically, through a menu, be able to select

what areas they want to focus on. And they are all risk-adjusted against everyone else. And the other critical thing is that you have got adjust across the country and that is something we do with our variables that we do make about 13 adjustments across the system to make everything sort of even.

So, the answer is yes, you can have a single but you can't just have that alone. There has got to be something with it that is actionable or else it is meaningless. And you know, I have heard that from, unfortunately, 300,000 people when we built the original model, tried to explain we are going to build sub-models and first we had to build the macro. And so part of it is also translating some of that to the public. They don't always sort of understand what we are trying to do and they sort of fix on one issue and not the overall scheme.

CO-CHAIR FLAMM: Yes. Okay, I have about six or seven people that are

already with cards up and that is the group that we are going to go through.

Alan?

DR. SPEIR: Thank you. I would like to echo the previous comments. And I would submit that while the presentation you gave is intellectually feasible, it is, practically, unrealistic. To have an all-or-none type of methodology is profoundly limiting and I think that it would open up that composite score to really be worthless.

Thank you.

CO-CHAIR FLAMM: Joseph.

MR. STEPHANSKY: When we were looking at the composite models, the thing that stands out is the weights to me. And I will bring this up again later in the meeting, I am very willing to have the cardiac surgeons and the interventionalists work together coming up with a set of weights to apply to the cardiac measure.

And over here when we are talking

about a hospital system, we have people within the system, the administration, at least, assigning those weights.

But once we move out of there, I think we are going to have some very difficult times trying to come up with any kind of objective way to come up with the weights.

So, I am starting right out as a dismal scientist, being very pessimistic, and saying no, we probably come up with these measures. We shall see.

CO-CHAIR FLAMM: Jack.

DR. NEEDLEMAN: I'm going to go back to the fact that we see measures in use. And I want to relate what we are seeing in use to this issue of composites.

So, we have seen in the CMS payment system, basically we divvy up the hospitals low, medium, high, in terms of cost, by some standardized measure of cost. We divvy up the hospitals low, medium, high, on some outcome like risk-adjusted mortality and

we sort of do a scatter plot. And when we have done these scatter plots, I have got some Ph.D. students who are busy working on surgical quality, so they are developing composite measures of complication rates among the surgeons and the hospital. The same sort of thing, low, medium, high; high, medium, low cost.

When you do that, what we usually see is a pretty smooth distribution across that chart. All nine of those cells are filled. And when we see that, we go to the upper left cell, low cost/high performance and we say congratulations, these are the winners. And we can get more for less, which is roughly the way CMS is doing the value-based payment.

Now, I can take those two scales and I can create some kind of weighting that basically tells me who is in that upper left cell. And in that case, I have just created a composite by combining the two scales. So, in some sense, I know I could do that. The

question is whether I get it.

So as I said that scatter plot is relatively smooth and I say there are clearly cases where I can get more for less. The knee in California. Right? We are fine.

What happens, one of the issues we have to deal with, the line items of that upper left cell is empty. The only way to get more is to pay more. We get more for more, less for less. And there, we need a different metric. And we have got examples of that in the cost manifest literature, which are ratios, answering Steve's question, where we got dollars for quality. And we can substitute quality for some other risk-adjusted scale for how much gain we are getting from the given service. And we could figure out how much more we have to pay to get those additional gains.

So again, yes, we can construct those measures in principle. We have done it. The question is, is this where we want to go

to help patients understand where their best choices are or to help payers think about how to pay to get value for the patients that they are paying for. And those, I think, are the open questions.

So, in theory, we can do this. Is it not clear that this is the right way to do it.

CO-CHAIR FLAMM: Because of the timing of this and our next discussion, I am going to -- Cathy, do you want to see? I was going to say Cathy and Dennis and did have -- okay.

DR. MAC LEAN: It is basically it is mathematically possible, but is it something you want?

CO-CHAIR FLAMM: Okay. Any other significant comments that you want to go ahead and add, Jeremiah?

DR. SCHUUR: Just sort of briefly. A comment would be that a cost quality composite, to me, doesn't necessarily create

an efficiency measure. We can have cost quality composite but we shouldn't label those all efficiency measures. And I think that is one of the things that is being done because we don't want to discuss cost in this country. So, it is more politically acceptable to discuss efficiency. That doesn't mean that we shouldn't aim to improve efficiency but a cost quality composite does not necessarily equal efficiency.

CO-CHAIR FLAMM: Herbert.

DR. WONG: I'll just take a minute to just emphasize the point that most folks have kind of recognized during this discussion, and that is that from a technical point of view, it is in fact feasible to create the composite measure of some sort. And I think the real issue here is the end user and whether or not it is beneficial for them.

So, let me give you an analogy because I just want through that, and it is my

daughter choosing a college and she made a decision last night. But you go through the normal process. You know the cost. It is tuition, room and board, et cetera. And then you have dimensions of quality. It is the prestige of the school. It is the acceptance rate. It is the return rate for the first year class and things of that nature. Many different dimensions. And she had a little spreadsheet that she looked at all of them.

Now, one could have came up with a composite measure of some metric quality divided by the known cost. But in the end, it was just this balancing act. Here is the trade-off. This has a good social environment but the academics is better here.

So, the question is really how useful is the composite measure to whoever the audience is.

CO-CHAIR FLAMM: Thank you. Sorry, I -- okay.

DR. WOZNIAK: That goes back to

NQF's sort of criteria generally around measures. So, other than usefulness, there is always reliability and feasibility.

And in terms of the composite, if I recall the framework, each of the individual components needs to be reliable and show that they make those sort of requirements. Correct?

DR. PACE: Actually, that is one of the things that changed in this latest guidance, where the focus is really on the reliability and validity of the composite score.

So, we don't require that the individual components individually meet the criteria or that they be individually endorsed. And I didn't want to get into all the specifics of the criteria, but ultimately what the last committee and the current guidance is, if you are creating this composite, it is going to be used as a composite. And that is where we need to see

the reliability and validity.

DR. WOZNIAK: I'm a little bit behind that NQF framework.

But the bottom line is the composite needs to be reliable and reliable to whom. So, it is really a combination of the use, the user, and the reliability.

CO-CHAIR FLAMM: Okay, so now we are going to shift gears and move into the next section on implications for efficiency measurement, as they translate into public and private programs, the MAP. So, Ashlie.

MS. WILBON: Thanks, everyone. I think one of the reasons we wanted to kind of keep the discussion going, I think we got a pretty good consensus from that last discussion on where that composite idea is going. So, but we wanted to make sure we had enough time to really talk about the next issue, which I think came up a lot in the first hour, kind of general discussion about the paper around audience and use cases.

So, a lot of the models that were found in the environmental scan were used for various purposes, depending on who the entity was that was identified in the table. We really wanted to kind of have this discussion with the group. We have identified kind of four use cases, if you will: quality improvement, public reporting, network design and tiering and pay for performance. And within those use cases, we wanted to have a discussion around what might kind of broadly -- again, we are going to have a more detailed discussion in terms of the application of the different models within the context of these different use cases in the breakout groups but before we got there, kind of wanted to have a discussion with the panel about broadly, what are considerations for measuring efficiency within these different applications or different use cases. And some of the examples that we list in the discussion guide on page four were some of the things that you guys

have already brought up, which is around audience, interpretability, and potentially the extent of scientific rigor.

So, I will leave it there and, hopefully -- I don't know if Taroon has anything to add. Okay, so we will open it for discussion and I will respond to any questions, if anyone has any.

MR. ROMM: So, and I think this is a perfect bridge of the last conversation to this. So, to Jack's comments and Jay's comments in the last conversation about the feasibility and then translating that to use cases.

We do this in Massachusetts and I currently create scatter plots of a variety of factors. So, a quote unquote composite of TME or of hospital operating efficiency. We wage index adjust it. We case mix adjust it. And we pivot it against things like readmissions, mortality, hospital process measures, a variety of other things.

And we use it for three of these four purposes. We use it for quality improvement in an investment program that I run. We use it for public reporting in our cost trends work. And we use it for pay for performance in that program as well.

The dial is almost immoveable. And so I think that one of the challenges that we find from -- immoveable. It is really challenging because the variation across providers on quality is so slim and the variation across providers on cost is almost entirely driven by price differential. And so, if you normalize prices, you don't see that variation.

And so I think that there is a sensitivity question here around those uses cases. While we have great interest but not -- when you go down the path of normalizing, any variation normalizes away and then providers just don't believe it. And so use for any of those three use cases, to date, has

been largely impractical, although we are trying.

CO-CHAIR FLAMM: Gary.

DR. YOUNG: Actually that was primary point, which is discrimination. And usually very often with measures, I find a lot of the quality measures that we have available probably do a pretty good job discriminating at the high end and low end. Then when you get into the middle, the information is not particularly helpful. And so, I think that is a very important consideration to what extent can these measures actually discriminate among providers. And again, not just on the high end or the low end but, as you get closer to the middle of the distribution.

CO-CHAIR FLAMM: Jeff.

DR. SILBER: I am kind of surprised to hear that after you did your adjustments you got no variation. We just -- there is a paper that we have just published looking at general surgery and orthopedics

across the country. A huge variation in outcomes and huge variation in payments for Medicare and costs. So, I am just -- I mean so that there are hospitals in all of those quadrants and huge differences. I am just surprised that you didn't see anything.

So, I am wondering if you, in some sense, by standardizing the prices you took away from any differences. But then I am surprised that you didn't get any difference in outcomes either.

MR. ROMM: So, just to respond to that, I think the balance is that the challenge that we face is that unless you overfit. I agree. I think that is the challenge that we face. We are overfitting everything. But unless you overfit -- yes, over-adjusting everything. And unless we do that, though, providers don't feel that it is appropriately applicable. And so that is the tension that we face.

And so, comparing a 73 percent

Medicaid mix hospital to a 2 percent Medicaid mix hospital, unless I adjust that, it is just not meaningful. But when I adjust it, the variation generally goes away.

So, we see substantial variation in cost. But that cost, when you normalize out the prices to utilization goes away. And so again, back to the sort of price disparity issues that we face in our market, that is a big driver.

CO-CHAIR FLAMM: Yes, Steven.

DR. ASCH: So, I would like to point out as probably many of us are thinking, the interim section between our last conversation and --

MS. WILBON: Can you come a little closer to your mike?

DR. ASCH: Sorry. I was getting a little too relaxed there, leaning back.

I would like to point out the intersection between, as many of us are thinking, between the last conversation and

this one, which is how do you aggregate measures. What is a composite measure? And what do you use it for?

So, one of the big knocks, as I think Peter was trying to point out on composite measures, is that as they get more and more composite, meaning more and more comprehensive in some sense, they get less and less actionable. People don't know what to do with them. But it depends on who those people are. And that is what these four use cases are pointing out.

So, it might be in quality improvement that the amount of composite measures that we have put into the numerator of an efficiency, well, the quality portion I should say of the efficiency measure would be smaller. Because in quality improvement, you want to narrow down and figure out what is it you are going to do fix cardiac surgery. What aspects of cardiac surgery are you going to fix?

For public reporting, as I think Joyce implied earlier, maybe I misinterpreted her comment, what do patients want? They want to know is it a good health plan. Right? And so you want bigger composites. And the numerator and the denominator have to match. So you have to have costs of the health plan and then some overall aggregate quality for a health plan, something that we all have reservations about but that is what patients would want.

And I can make the same arguments for the other two use cases at varying levels. So, I think it is important in this paper, as the paper already begins to do that, I believe, to make sure that if we are talking about composition, we are not just talking about the composition of cost and quality, so the ratio, but we are also talking about the scope of, meaning the composition of both the numerator and the denominator, making sure they match.

CO-CHAIR FLAMM: Andy, you wanted to comment?

DR. RYAN: Thanks, Carole. So, I had two comments; one about the issue of discrimination. And when we did the environmental scan, I think that, particularly on the private payers, there was the same concern, even if it wasn't stated explicitly, that Gary raised that on the margin, these measures might not really discriminate very well with providers but maybe it can say these are better. In other words, this is the group we want to note for distinction and say these are the best. These ones we think are not as good.

And so, I think that was kind of a concession in how kind of the tiering is done with a lot of these in practice and how distinction is kind of noted that maybe that broad groups can be identified but incremental differences in these measures probably don't mean that much. And I think that concern is,

in some ways, reflected in how the measures are actually being used in the private side.

I wanted to also comment on something that Herb said about his daughter making the decision for colleges. And you know it is a very interesting case study to me. So, she looked at these factors and said I come up with an answer I think this is best. But did she make the decision that is going to optimize her happiness? Because now we know from all like the behavioral economic stuff that people make decisions that are systematically wrong. And there is probably --

(Laughter.)

DR. RYAN: If she had just chosen, if her simple metric was distance to the closest beach that probably would have worked.

And so there is a similar analogy with how we are doing this in healthcare. Because you know patients could look at a series of measures and say I think this one is

the one that matters to me and matters for my outcomes. But that might not actually be right.

So again, it is a question of how we want to think about that kind of architecture of choice that might be, I think, a contribution that this group could make.

CO-CHAIR FLAMM: Good points. Timothy.

DR. LOWE: I think Iyah hit the nail on the head. And that is, a lot of what drives this is buy-in. The frustration that I have for developing measures is that I decide to do measures and my primary focus is what is most clinically meaningful. What areas can we say that a clinician can actually make a difference in? When I think apply that to data analysis, I find that it discriminates very poorly between them.

If I come at it from a data mining perspective, I can find things that discriminate very well between providers but

aren't particularly clinically meaningful. Even if I am able to come up with some of them, as soon as I send that out to our members, I get the have you risk-adjusted these.

So, when you risk-adjust them, then sometimes the differences become very small. You find out the same thing. It depends on how you risk-adjust and whether we are over-fitting our models because some things we just don't know.

So, the question is, for me, always, what is going to motivate people to change their behavior. And that changes then how we approach the measurement process. These may be great measures but if these don't result in the changing of provider behavior to make it better for patients, then why are we measuring those things? Why do I spend my time doing this if I am going to get questioned or people are going to be suspicious that does this really work or is

this really fair? It has to be fair and it has to be workable.

CO-CHAIR FLAMM: Okay, just to put a little context on this part of the discussion, we will take these comments for those who have cards up right now but then we are going to shift into another part of this discussion.

So, I have Dennis next.

DR. SCANLON: So, I think the way this is organized up here for these four different purposes but I actually would suggest that these are very much related and there might be some utility in thinking about how they are related from the perspective of those that might implement these types of things.

So, take quality improvement, for example. I may provide high quality very efficiently. I may produce high quality very efficiently as a provider and may not need to sort of engage in efficiency-based quality

improvement if my market share is something I can maintain and if there are market forces that doesn't cause me to sort of address the efficiency issue. So, that sort of illustrates how sort of the reason why I might want to engage in efficiency-based quality improvement is largely due to some of these other things, like payment reform, which maybe we will call pay for performance, or public reporting, where someone has a stake in the game financially that sort of forces me to do so.

So, I think that it is useful for thinking about these as separate but I do believe that they are interrelated and nested. And again, conceptually, there might be some value in really thinking about how these things relate to each other.

And ultimately, I would suspect, I am not sure that I am correct here, that at end of the day, and this gets back to sort of the definitions that we have been talking

about, really what might be the kind of common element here is value. What this sort of converges to is a situation where society, individuals, employers, whoever, want to sort of get the appropriate level or the best level of quality for the minimum amount of expenditure.

Now, providers might not have an interest in sort of making that happen, unless the market forces sort of encourage them to do so. And consumers may not have the tools to sort of help them make that happen.

So, I think there is some utility in thinking about these separately but they are related. But I guess at the need of the day, in my mind, is value kind of this overarching thing that really is the reason for talking about efficiency.

So, should the question be what are the considerations for measuring value and efficiency as kind of being a component of that? And sort of what folks might do with

this is all around sort of the framework of trying to get at value.

CO-CHAIR FLAMM: Okay, Larry.

MR. BECKER: Thank you. So, I agree with Tim. And I think these are signals for providers to make changes and for consumers or patients to make choices. And I am really troubled by the notion that there is one right, that one of these measures embedded in it, if you have a single measure, is somebody else's point of view.

And that what I think this is about is a series of choices across a series of dimensions that the biggest computer in the world couldn't solve for. And so, each person, given their circumstance, their genetics, their family history, wherever they are in life, makes a series of choices. And for us to try to get it into a single measure, I'm sorry, I think it is chasing a unicorn.

CO-CHAIR FLAMM: Iyah.

MR. ROMM: I guess I somewhat

agree and I also think though that, especially for those who are represented around the table, that part of where we really need to understand where we are going, especially with the third use case, because I think tiering, in some ways, is the greatest opportunity here, to figure out what the underlying use cases even of tiering are.

Because what we are finding in Massachusetts, specifically, is that largely quality is being pushed aside in tiering and tiering is happening almost entirely based upon TME. And in fact in many cases, folks have noted that that is a conscious choice. So, we are undergoing an interesting conversation right now, trying to decide if we are going to standardize quality measures across plans for tiering. And uniformly, the answer is no, both from the providers and from the plans, which is very interesting.

And so I think that back to this question of is chasing uniformity a unicorn,

I don't think it is. I think we all want some standardization. And frankly, I think that for a consumer to understand what that means is essential. We have to have uniformity.

But I think that the idea that we are going to boil that into a single measure or a single set of measures I tend to agree with Larry. I just don't think that is what the future holds.

DR. BURSTIN: Well, why don't they want it?

MR. ROMM: Because I don't think that most consumers, from my perspective, are looking at cost and quality as being the same thing. I think they are looking at as being pivots of a given experience. And so, rolling it up into one, I think it is an individualized variation. So, I think that part of the challenge is that if we look at this, this definition of sort of self-determination of what value means, you can't say for low SES versus high SES consumers it

is one thing. I think it is an individualized choice.

DR. BURSTIN: I think you said that the plans and the providers didn't want uniformity.

MR. ROMM: I'm sorry. They do not want uniformity because it is a market game. So, back to the comment earlier, everybody wants to be able to be able to build their incentives in their own way and they don't want to treat providers the same way. And so the interesting thing is that providers don't want uniformity either. The plans, I get it but the providers don't either.

PARTICIPANT: To avoid black box, you make choices.

MR. ROMM: Because it is still all a negotiation, right? I mean even underlying tiering, it is still a negotiation.

CO-CHAIR FLAMM: So, this conversation is clearly getting on a roll, which is a very good thing for the breakout

groups. So, largely, we are going to continue this discussion in the break out groups. I think, Alan, you had one last comment and then we are going to move into the section about the MAP discussion.

DR. SPEIR: I want to discuss examples of what we are talking about about behavioral change, about both physicians, patients, and insurers. In the original Medicare demonstration programs in the last of the '90s that were tasked with quality and reducing costs, only one of the hospitals actually was able to achieve that. That was Saint Joe's in Atlanta. And it was directly tied into incentives to the surgeons themselves.

So, unless the physicians were incentivized, then they weren't going to change their behavior. And this has the practical implication now because over 80 percent of physicians are now employed by their healthcare system. And if you look in

Virginia at the 12 practices in 18 hospitals, their entire reimbursement is predicated on RVUs, volumes and they are not tied at all to achievement of any quality measures, which I think is a difficult subject.

Secondly, in terms of motivating patients, it has been my experience and those of my colleagues that patients view their hospitals in their locale as providing excellent care. And getting them to move from one locale to 12 miles away to a nationally-based hospital with excellent outcomes doesn't happen. You would think that it does but I think that is an opportunity for groups like this to really focus on efficiency and what outcomes really mean.

And that goes to the third point around insurers is that there was an extraordinary amount of effort in the last 12 years around Centers of Excellence. But yet while those designations were made, particularly around cardiac surgery, the

patients were never challenged. They were never given the directive that they would pay lower deductibles to move from one center to another around outcomes. So, what difference did it make to make those designations?

But I think that the more focus we have here on efficiency, on outcomes, is a huge window for education for all of us.

CO-CHAIR FLAMM: Yes, and the trend towards increasing those incentives and benefit designs is something that has been picking at more recently but early on for certain.

Okay, do you have a question, Dennis?

DR. SCANLON: Yes, it is just a question. I guess I am wondering why regulation isn't up here. And is that sort of implicitly subsumed in one of these other categories? Because I think a lot of -- you know some of the discussion here is around these trade-offs and, presumably, regulators

in terms of establishing minimum quality thresholds, making sure that the cost dimension doesn't sort of go too far down the track for populations that are being covered.

So, I guess a question is whether uses for regulation is something that should be thought of separately or whether that is subsumed I any one of these categories.

CO-CHAIR FLAMM: Okay.

CO-CHAIR DUBOW: Well, pay for performance incorporates regulations. I think the pay for performance assumes that there is a regulatory process and we, certainly in the context of the MAP discussion with the MAP considerations, where MAP is making recommendations to the Secretary about the measures that should be used or over 28 federal programs, there is a regulatory underlay, undergird what the Secretary is going to be using it for. And her authority is regulatory, usually, or based on statute. But it is reflected in regulation. So, I

think it is sort of implicit.

I think you probably could think about regulatory underlying some of the public reporting. It depends on who the payer is. If it is a public payer, there is probably a regulatory structure. And even in the private side there are regulations and statutes that govern some of the things that happen.

So, I think it is implicit.

CO-CHAIR FLAMM: Okay. So, shifting into the implications of MAP.

MR. AMIN: Yes, so I am going to actually ask Erin to take the lead on this section. But the basic question here is to think about the question of selection, of how the models and the particular questions of how measures are selected and how that interacts with the models is really the section, this next section. But anyhow, I will turn it over.

MS. O'ROURKE: Thanks, Taroon. So, I am one of the NQF staff supporting the

Measures Application Partnership and Joyce is a member of the Coordinating Committee overseeing this work. So, we wanted to bring this to you as a way that NQF will be advancing the work of this panel and making this not purely an academic exercise but using the guidance of this committee for real world applications.

To give you a little bit of background, MAP is a public-private partnership convened by NQF where we provide input to the Secretary of Health and Human Services on the selection of measures for public reporting, pay for performance, including a number of programs intended to measure efficiency.

MAP's goals are to identify the best available measures for use in specific programs, to provide input to HHS on measures for specific uses, -- as Joyce mentioned, we provide input on over 20 federal programs -- and to encourage alignment of public and

private sector efforts.

So, I have noted a number of these programs are intended to be focused on efficiency. The main way MAP provides input is through our annual pre-rulemaking process, where we receive a list of about 500 measures for these 20 programs and make specific recommendations on whether or not MAP supports HHS proposing these measures for use in their programs.

So, we are really looking for guidance from this panel to help us refine our decision-making for these efficiency programs. I think I was struck in the paper noting that the importance of the measures you put into the model is really tied to whether or not you are actually measuring efficiency. Otherwise, are you just measuring unrelated cost and quality outputs?

So, these are the current MAP measure selection criteria. I don't want to read them out to you in too much detail. I

should note that we consider these at the program set. These are not criteria meant to judge an individual model. But what we are looking for here are there additional measure selection criteria, if you will, that we should be thinking about for efficiency programs and what guidance would this panel have to the MAP as we make recommendations on use of measures in some of these programs, such as the value-based purchasing, the Medicare Shared Savings Program, or the physician value-based modifier.

So, with that, I will turn it back to Ashlie and Taroon for discussion.

MS. WILBON: Yes, I will just add a couple of things. One, in the break we had a discussion with Joyce and she pointed out that for the purposes of the work that NQF does, obviously, we were looking for operational guidance for the measure applications partnership but that broadly, that this type of criteria could be used by

other entities as well, who are looking to select measures and into programs for their own purposes.

So, we have framed it in the discussion guide and for the purposes of this panel based on kind of our own selfish needs for potentially additional guidance for the measure application partnership but also wanted to include the broader context of these type of guidance could be used for other entities as well. So, I just wanted to include that.

And I wanted to also mention that we are going to, I know we moved away from the measure selection criteria on this slide but in your discussion guide, it is also listed and we are going to have you guys, the next probably 25 minutes of our discussion will be focused around the questions in the box on the top of page five, which include should the selection of measures for a model or approach vary depending upon its use case and whether

or not the selection of measures depends on the purpose of the model or does it not matter which approach we are going to use to measure efficiency and that the selection of measures is kind of agnostic from the actual approach used by the model.

And whether or not there are other models that we should be considering that are not included in the paper that Andy and Chris found in their environmental scan that we should be considering as a part of our discussion for the next section.

Following this, we are going to have Andy and Chris really go into some more detailed description of what they found in their environmental scan in a more detailed description but wanted to get some overarching ideas about that as well.

So, with that, I will open it up to the group. Andy and Chris, do you have anything? Okay, you looked like you were getting ready to comment.

DR. TOMPKINS: Well, I was actually going to pose that question that you just posed. Meaning, since you posed it, I want to emphasize it. Before we get into breakout sessions, does somebody have to put forward now a model that doesn't fit in the taxonomy in the listing that is in the background paper? Because if the answer is yes, then as all the breakout sessions consider the models, they should be an inclusive list and not just the ones artificially in that case restricted to what was in the background draft.

If the answer is no, that's fine. But since you asked it, I thought, we don't want to have some undisclosed model.

CO-CHAIR FLAMM: Go ahead, Donald.

DR. LIKOSKY: I just have a quick question. Have we come to terms with how we evaluate the models, what systematic way we evaluate them?

I mean in part I am still trying

to address the questions posed on the top of page five and going back to how do we evaluate it. Have we come to terms about a framework or is that really the purpose of --

CO-CHAIR FLAMM: In the breakouts, I think we are going to talk about pros and cons and the different kinds of issues around applying them in the use cases that we are focusing on. So, I think that is going to be fleshed out in terms of evaluation aspects.

DR. LIKOSKY: Okay.

DR. TOMPKINS: Well, I don't want to necessarily increase the scope of the work but I will make a suggestion that the breakout sessions don't have to be strictly along the lines of okay, Model 1, Model 2, Model 3, as if that is necessarily the most productive way to do it. Although, contradict me if that is exactly what you want people to do.

Because it might more liberating and more useful to have people say oh, I'm sitting in the public reporting group. What

of all the issues swirling around in the room are the most important ones to remember in any step of this, whether it is one of the measures, process versus outcome, whether it is the reliability issues, things like that.

And then to the extent that that MAPs better to one model than another, then that could be an intermediate finding in the group. But otherwise, I think an enumeration of the use cases and taking that perspective and the interesting contrast would be if people come back with either a different mapping to the models or a different sense of the way in which the construction of the linking process would ensue should be noted.

CO-CHAIR FLAMM: Okay, all these cards appeared at once in my mind as I was looking. I am just going to go in order around here. Go ahead, Iyah.

MR. ROMM: Sure. I'm not going to be very helpful. I am going to introduce a question and not an answer, which is to say

one of the things I was struck by in reviewing these is the extent to which most of the models that are out there, and this has certainly been our experience, adhere to one of the models of your taxonomy, Chris and Andy, and I don't have a suggestion of another one that doesn't MAP there.

But I am going to pose the question of should the model drive the measures or should the measures drive the model? That is, I think that all of this is sort of framed around the idea that within a given program we want to choose a model and then find measures that fit.

I am not sure, given the state of measurement of efficiency that that is the right approach.

MS. WILBON: That is, in fact, one of our questions for you.

CO-CHAIR FLAMM: Cathy.

DR. MAC LEAN: So, I 'm just going to echo that and I was going to ask the

question is the purpose of the discussion to discuss these different models and understand how things work or don't. Because I have to say as I went through these models, there are a couple of situations that I really couldn't quite figure out where they fit in. So, bundled payments and part of the issue with the bundled payment is you are negotiating a rate with the provider. Right? So, right there, it is all about the negotiation. And so, how does that fit in?

Or the reference-based pricing example that we have talked about a few times, how does that fit into these models?

So, my question to you is, as we go to do this, what is our task? Are we using these models to raise important discussion points or what is it that we are tasked to do?

MS. WILBON: Are you asking specifically about the breakout groups or about the -- I just want to make sure I can address your question appropriately.

DR. MAC LEAN: Both the breakout groups and what is it you want us to do with these models, either now or in the breakout groups. What is the purpose of these models? Is it to engender discussion or is the purpose of our work to try to fit things into these models?

MS. WILBON: So, part of the scope of the work and I will open it up to Taroon as well to discuss, was to do an environmental scan of what is currently out there. So, these models are a reflection of what Chris and Andy kind of found of what people are actually doing out there now. And that is reflected in the table and from the different entities that they have found.

So, it is not that we are trying to -- this is kind of what is out there. We are not necessarily trying to fit anything to the model. I think it was intended to be kind of an exhaustive, to the extent that they could find, an exhaustive list of kind of this

is what people are doing out there. And then for us, for this panel to provide some guidance on what is the best application of these models, is the way they are being used now most appropriate and is there some additional guidance we can provide, based on the context of those models, how measures should actually be -- cost and quality measures should actually be brought into those models in order to measures efficiency.

I don't know if that is helpful. Maybe Taroon can be more articulate than me.

MR. AMIN: I will try. I think you have characterized it pretty well. I think -- so, let me just describe the current state.

So, we have these cost and quality measures that exist sort of independently of one another and then we have programs that are sort of using them. And there are various different ways that these programs are using them and the purpose of the environmental scan

was to describe the models in which the programs are using them.

Now, the basic question that is being asked that we are interested in understanding is that, first of all, the way that the programs are designed, meaning the use of the program, whether it is for public reporting, for payment applications, we want to understand in more detail how the model selection and the use case are related or not. That is the first question, which is why maybe the use case is driving the model design or not. Just understanding and characterizing that relationship.

The second is to understand right now we are selecting measures for programs, based on its use case but there is also this intermediary where you have, there is various different models that which programs are sort of using in their program design.

And it may seem like just selecting the measures without really thinking

about the way that the model is constructed might not be the best approach. That you might actually need to think about the program, the model, and then selecting the measures for the model. And so that is the second dimension that we want to explore.

DR. RYAN: Catherine, so with respect to where does bundled payment fit in, if you look at a number of the programs we have identified use some kind of episode-based measure of cost. So, some kind of like looking over an extended based on some index event and assessing cost within some bundle of services over some specified period. And that is how cost is specified.

And then so, then that program might have some other way of looking at how quality is specified and then some model for combining the cost and quality measures together.

So, the bundled payment part is just kind of how cost is specified and the

model is how that specification of cost is then subsequently combined with quality to give some measure of efficiency. And also with reference pricing, the one real example I think is CalPERS. And there is no explicit integration of quality information right now with CalPERS.

Jamie talked to our group. It was like well, you know we think these are high or low quality. We think we got people who provide reasonably good quality kneed replacements or whatever who can do this for $15,000. And so that is our price. We think we have some people who are doing that.

And so there wasn't like an explicit way to integrate that cost and quality signal in that program. So, that was left out of our table.

So, I don't know if that is helpful but that is kind of the rationale for how we organized this.

DR. MAC LEAN: Well, just for the

CalPERS, Anthem administered that. And there were quality criteria. Now, they were pretty minimal quality criteria like the hospital had to be accredited. There had to be some minimum volume of procedures. So, there were quality criteria. They were quite minimal, I will admit. But they were there.

So, and I guess maybe in the side-by-side model but I don't know that it even matters. I'm just not sure where we are going with these models. Are we trying to figure out how to fit these into the models or are we just using this as a framework to kind of move our discussion?

CO-CHAIR FLAMM: And I don't know if this helps but I think Chris' comments earlier about this idea of maybe using our breakout groups to talk about from this perspective of this use case, what are the requirements that I am trying to accomplish?

It might actually help filter through the models because each model had

different implications in what it produces. They don't produce all the same things. And maybe there is a different applicability or fit between each of the models and the use cases. We sort of popped that through.

Craig.

DR. WOZNIAK: I guess backing up there was a question about are there other models that we ought to consider or that are not listed. And I will go back to my point I made earlier and Jack said they are in there. But I am not sure if measures that are developed with the cost and the quality measure developed simultaneously are included in this table.

DR. NEEDLEMAN: Nobody is paying them. But that is the difference.

DR. WOZNIAK: Well, that is a different issue, right. Are there are other models that ought to be -- the question was are there other models that ought to be considered? And I am speaking more on the

physician side than on the hospital side but, take for example, on the physician side you have quality measures that have very well-defined specifications that maybe are pulled from EHRs, PCPIMA measures, for example.

The cost measures may be totally different in terms of the time frames, even the patient population so that you may have different populations that are being used to measure the quality than you are to measure the cost. You have an aggregate cost measure, for example.

Or again, you could have an episode of care with a certain time frame and a certain population of patients. That is not the same population of patients that goes into the quality measure. So, you have got sort of an odd mismatch of the denominators, the numerators, the levels of analyses and how you aggregate it up. So, that is what I meant in terms of having measures of cost and quality developed simultaneously. So, measurement

periods match, populations match, the unit of analysis match. The components that go in are consistent across the elements of these two.

And again, it is probably there is more of a void on the physician side, possibly. Maybe it is less applicable on the hospital side but I don't see those kinds of measures being listed in the table. I mean, clearly, side-by-side means there a measure of cost and there is a measure of quality and maybe some of the patients are the same. Maybe some of the docs, the sites are the same, but they are really not the same denominators.

CO-CHAIR FLAMM: Okay, Dennis.

DR. SCANLON: So, one of the things I thought about in looking at the table and the paper, which I agree was helpful to sort of put some of these programs or things that are happening around the country on paper, is that while we sort of identify which model among the category that the authors sort

of put in play, these programs are all for different purposes. And there may be some value in trying to sort of organize them by purpose or primary purpose and then see if there is some convergence around some of the model. So, just as an illustration, the Medicare Shared Savings and Pioneer Program, I sort of view that as the purpose is to incentivize innovation on the delivery side to take care of patient populations for a cost that is lower than traditionally we have paid and to sort of develop a financial incentive so there is shared savings around doing so.

So that really, the purpose there is really around delivery system innovation for purposes of accomplishing a broader societal goal, which is to reduce healthcare trend.

Contrast that with sort of number 23, which is Castlight. And you know what I know about Castlight is that is really a tool for consumers to make decisions, largely in

kind of a high deductible plan context at a point of service or something, completely different from the Medicare shared savings program.

Health Partners, what I know about that program is a lot of the reason for Health Partners sort of developing their tools was to become more efficient internally in their delivery system and have a way in kind of an apples to apples fashion to kind of compare provider to provider within their own integrated delivery system, where they have a high degree of control over providers within that organization. So, there was a real internal purpose to that.

And then if you take a look at others like episode payment approaches, those may be for purposes of setting reference prices or developing tiered networks, which might be sort of objectives of health plans or payers.

So, one thing that I think might

be helpful is to really categorize these things and then look and see if you look at the sort of innovation efficiency of the delivery system ones, do certain models seem to fit where there are a subset of models that seem to apply. I think that would be sort of a good step.

But I also sort of want to say that in terms of alternatives, I am not sure this is an alternative model. If you really want to think outside the box, it goes back to our discussion at the very beginning of today of what is sort of the goal or purpose of this is. If the goal or the purpose is to sort of create a more efficient health care system without sacrificing quality or perhaps even improving quality, there could be other sort of approaches to this. I mean you might sort of take a measure which is not condition-specific like of them are or not episode-specific but you might look at investment of organizations and what they are doing to

address things like waste or to redesign their delivery system. It might be more of a structural measure in the sort of Donabedian traditional kind of approach of thinking about measurement.

But it is sort of more what is being done in the name of sort of reducing waste or redesigning the delivery system or creating innovation, which is entirely different than a lot of what we are seeing currently, which is clinical or disease-specific type measures.

So, that is a little bit outside of the box but I guess it gets back to why are we doing this to begin with.

CO-CHAIR FLAMM: So, we have run a little bit over our time. Let's take about five more minutes for this discussion, then we will move into the public comment. So, we have a lineup of folks that need to speak. I had Matthew next, and then Jack, and then I will circle around.

DR. ROUSCULP: Thanks. And again, I am following Dennis. He laid out a lot of very great points. So, I am not going to repeat all that.

I think the one area that I am hearing from you at NQF is you are saying we have to be pragmatic here. You are saying that you get a lot of quality measures and you are starting to get cost measures. And CMS is coming to you or whatever the organization and it is saying hey, we have legislation that is dictated. We have to talk about efficiency. It sounds to me, then, because of that, that you are asking that when cost measures are starting to come in is that you don't want to just have cost measures but you want to say how do those link back to the quality measures that already exist. I think that is what -- at least that is what I have taken. I could be completely wrong. I have been wrong before many times.

So, from that, I guess what I am

trying to think about as far as the models themselves, it is not necessarily are these the penultimate models to move forward but it is saying within what you are being asked to take on at NQF, what are the ways that you can have guidance from us that you can look at whatever these cost measures come by that you can turn around and ask that before we accept your cost measures, you must show us how you are linking this to quality.

And so to me, I want to make sure either A, that is correct and then B, it is not ever going to be one of these models is going to the penultimate. What is going to happen is these models are going to be again dictated by what Dennis was bringing up, which is saying whatever the end point is it about decreasing costs? Is it more about increasing performance? Is it more about increasing the quality of care or decreasing the safety signals that you are starting to see? I think that is where we are going to kind of winnow

down and we can kind of help with you on some of those elements and say which of these models can be helpful or where are the ones that aren't here.

CO-CHAIR FLAMM: Jack.

DR. NEEDLEMAN: I think there are four things and these all relate to issues we have talking about here.

First of all, we have got a lot of models. Some are in use and some aren't. And I will come back to the ones that are in use in a moment. But things like data envelopment, stochastic frontiers, the regressions models, not being used in any of the programs we see. They are very different in concept. They are out there in the economic literature. I played around with them. They are all feasible to do but nobody has gone that way.

So, some are in use, some are not. The ones that are in use are all a variation of the side-by-side. Fundamentally, you know,

side-by-side, hurdle, conditional, unconditional, they all basically say for the cost -- they are all some variation of the scatter plot, throw them up on the cost dimension, throw them up on the outcome dimension, and then take your eraser and decide where we are going to start erasing and saying these folks are in and these folks are out or these folks are getting paid well and these folks aren't.

So, they are all basically variations of the side-by-side and all the action in those are really around what determines where you get ranked or what you get scored on each of the individual dimensions. And those raise all the issues that Greg raised.

So, in reviewing any of those measures, you come back to where did they get the data from? Is it consistent? How are we weighting things? How much are we pulling things in a Bayesian way? And on, and on, and

on. Lots of very specific issues.

But those four measures, the ones that are in use are basically the same measure. And so, you have got that issue.

I think the question becomes a question of robustness and relative performance. When you begin changing these things, when you begin changing how Joe's daughter weights closeness to the beach versus academic rigor, do you come up with the same ranking or a different ranking? And we haven't talked about the robustness of these measures as a critical component of deciding how much confidence should I have in them.

And then we get to the interpretability and that may affect -- you know if they all come to the same answer, then the interpretability determines which one you pick. But if they come to different answers, we have got to spend a lot of time thinking about getting which answer we believe the most.

CO-CHAIR FLAMM: Gary and Jeff. Steve, you want to make a quick comment?

DR. ASCH: It's just a point, a very quick point of clarification.

So, the VA actually is using stochastic frontier models for efficiency. It poses all sorts of application problems, which I would be happy to talk to you about.

CO-CHAIR FLAMM: Okay. Gary and then Jeff.

DR. YOUNG: Well, I guess we are all struggling with sort of the same issue, same concern about trying to think about how we might have a meaningful discussion about these models and trying to think about what again is the large goals. And again, going back to this morning's discussion, the earlier discussion about potentially the need for some sort of overarching framework because placing those models in context is going to be hard without doing that and having a meaningful discussion. I mean, many people, for example,

in their comments, have talked about the importance of the actionability of the models and the measures but from a consumer standpoint, they may not care about actionability. They just want to be able to make a choice. So, whether or not the provider improves or not is not terribly important.

Other comments have talked about whether having a single measure or a single score is feasible and useful. But, obviously, it doesn't have to be only a single score. There is also drill-down opportunities. So, is that something that needs to be considered as an important consideration. And then, of course, there is stakeholders. So, going back to the example of choosing a college and beach versus academic rigor, parents and kids tend to have very different preference weighted considerations around those kinds of dimensions.

So, I think that needs to all be

considered as part of a larger framework to have a meaningful discussion about these different models.

CO-CHAIR FLAMM: Jeff?

DR. SILBER: Yes, to answer Christopher's question, the list of models is not complete. And I think, in part, you could have given more weight to the side-by-side in terms of there are variations on the side-by-side model that have been not explicitly made clear in the document. And so that, basically, for the first models before you get the side-by-side, we are talking about efficiency in one way or another that is incorporated into the model. And side-by-side splits those out but there are so many ways to do that, that I think you have underplayed that in the document.

So, that would be the big problem that I see with the document and also in terms of answering your question, we have to also think about how to present side-by-side in

ways that are most useful to the consumer.

And I don't think efficiency is something that the consumer is necessarily interested in, the consumer being, facing north, the patient. They want to know price and they want to know quality. Maybe they don't want to know quality but we hope that at some point they will.

CO-CHAIR FLAMM: All right. Thank you, everybody. This was a really wonderful discussion. I think it sets us up well for the breakout sessions that we are going to have after lunch.

At this time, we would like to open it up for public comment. So, we will do two things. Operator, if you could let us know if there is anybody on the phone who would like to make a comment first.

OPERATOR: If you would like to make a public comment, please press *, then the number 1.

At this time, there are no public

comments.

CO-CHAIR FLAMM: None? Okay, we do have some guests in the room. Public members, do you have anything you would like to add?

MR. HAIDER: Shall I speak from here? Can everybody hear me?

Thank you very much. Thank you very much for providing an opportunity for the public to give comments in such an esteemed group.

So, I am a technology entrepreneur. I have left my tech world and moving to healthcare to start a company to help patients. And I have a personal story behind that that is driving that.

So, I have to analogies that I hope will help the group from a tech perspective. So, the first analogy is the cell phone market. And I have kind of mentioned to two people while I was talking about it. So, the cell phone market is

continuously giving us higher quality with more features and reducing cost. Right? Now, if you look at your cell phone bill 30 years ago, how many of us had a cell phone bill? Zero. But you look at your cell phone bill today, I am quite astonished how much money I spend on it and I am happy to do it. So, even though I am spending more money, I am happy to do it because it is providing a value to me.

So, what I mean by that, where I am going is that tech markets are extremely efficient. If you look at the cell phone guys, they have just taken the cell phone that cost $20,000 at one point and now driven it to $200. It was only available to the elite. Now, it is available to everybody. And everybody is happy paying more. The market has grown.

And so, the reason why that tech market or many markets, as you go into the economics of it, is because of the ability for choice. And that is what I would urge the

group is to look at -- I urge the group to look at the efficiency in the context of the market. And as part of the white paper, add the fact that we are restricting patient choices is a problem.

That being said, I want to bring one more analogy for constraint. In the tech market, if I choose to buy Apple, I am constraining myself. I am going to get just a certain type of power adapter and monitor and stuff. I can choose to go with Samsung and I will get maybe a different type of Android-type ecosystem. But again, that was my choice. So, based upon my choice, I might enter into certain constraints.

So, in the healthcare market you can think I may enter into different plans. So, that is the first point.

The second point I would like to make is on feedback on quality measurement and models. So, open markets have quality measurements. They are determined by

prioritized by my, as a consumer, individual values.

And the analogy of going to college, I want to choose. Do I want to go to a place where there is a beach or do I want to go where there is no beach and it is all studies, or do I want to try to get both? And the way that is done today is that those values are communicated to me through brands, Consumer Reports, J.D. Powers, Yelp, U.S. News. All of these provide different values. And I may read Consumer Reports but I don't look at Yelp. I may look at Yelp but not J.D. Powers because that is the way I like reading reviews.

And I think the industry groups and government's role is very important, and especially this industry, to enforce transparency and meeting safety guidelines like CE Mark and so forth. So, I, as a consumer, can make safe choices. But still, I am allowed to make that choice.

And then in this regard, I urge the group to look at metrics that need to be reported that will help people make choices based on their wide ranges of values.

And the analogy, one more analogy, there is the food label. That hey, I want to choose how much calories I am taking, how much protein I am taking, and how much -- but I can choose whatever I want as an individual.

And in final, I have one point on the culture. When I go to my doctor, I feel like I am talking to my dad when I was ten-years-old. And Dad says I am going to make this choice for you. And here is my response to my dad, I am going to say, Dad, please tell me what you think is important. Give me your experience but let me make my choice because I have to live with it.

So, those are my points.

CO-CHAIR FLAMM: Thank you. Are there any other comments? Okay, thank you.

Well, with that, then we are going

to move to lunch. We would like to allow a half an hour for lunch. So, if we could start to gather back here right around one o'clock, we will get started shortly there with the session at one o'clock.

All right. Thank you, everyone.

(Whereupon, at 12:31 p.m., a lunch recess was taken.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

(1:16 p.m.)

CO-CHAIR DUBOW: Can we reconvene, please? We have a busy afternoon. And we also have an important announcement, Erin, before you begin.

Could everybody please at least be silent so we can year the important announcement that Herbert wants to make?

DR. WONG: So, many folks have been asking me what the conclusion of my daughter's decision was. So, I will say that she did create a spreadsheet. There were all sorts of quality dimensions put on this. None of them really had distance to the beach on it but you know, it is possible that she was keeping that aside, absolutely. So, she had other things such as school rank, size of the population, the percentage of first-year that comes back to the school as an indication, enrollment dollars or endowment dollars for the university. So, it was a little bit

sophisticated. I am fairly certain that the distance to the beach was not in her choice set because the school that she selected must be at least 200 miles away from a beach.

So, in the end, she selected Colgate University, Hamilton, New York, in Central New York, a very isolated place. And Mom and Dad are happy with her selection. So, there we go.

PARTICIPANT: There is skiing nearby.

DR. WONG: There is probably cross skiing nearby and things like that, absolutely.

CO-CHAIR DUBOW: Okay, Erin.

MS. O'ROURKE: So, we wanted to spend some time now to let Chris and Andy really tell us the details of the approach they took to the environmental scan and walk us through each of the models and what they entail, and then give the committee some time to weigh in on both the approach to the scan

and the models, generally.

DR. RYAN: Great. Thank you, Erin. So, I am going to start and take you through the methods and results and just some kind of high level summary thoughts about kind of what it boils down for the group. And then I will let Chris comment on this.

So, what we did was we did a couple -- had a couple approaches to the literature search. We did search the PubMed databases for English languages. Articles in the last 25 years that had -- you know we had a variety of search terms we used and quality measurement, cost efficiency. And then we tracked down articles that use these.

So, I want to say that as Herb mentioned before, AHRQ contracted -- did a big contract with RAND a couple of years ago to summarize the state of efficiency measurement. And so that was a nice kind of input to this process, but we didn't try to duplicate their systematic review.

So, we did that. We pulled articles that way. And actually you know we emailed everyone on the committee. I apologize if we missed any of you but I don't think we did. And we got a lot of great feedback that way. And a lot of really key materials that kind of led us in different directions. So, a number of people were very helpful for that. And that helped us identify programs.

So, this led us to identifying 30 plus programs. And then we kind of dug through. We took information that community members gave us but, where possible, you try to find supporting information on the websites and contacted people to follow up with specific program information.

So, we compiled this list of programs that are using, that are linking measures of cost and measures of quality. And again, this is not -- the perspective we had were the cost was not on the kind of provider side or production costs. It was costs for

the payer, the purchaser, or the patients, since that was our perspective in identifying programs.

So we identified about 30 programs. We filled in information on kind of the what kind of provider was being profiled, was it hospitals, was it physicians, was it healthcare systems. How was quality specified? How was cost specified? And then what we did was we went through and said what are all these different -- how can we come up with a taxonomy of ways for combining quality and cost measures. And after kind of reviewing all these programs and kind of iterating through, we came up with these seven mutually exclusive models that either programs have used or a couple of these have just been proposed by researchers but they met our other criteria that they were designed to profile individual providers on the basis of efficiency, based on cost and quality performance.

So, and let me just I will go through and kind of describe what these models, the key features of these models and then come back to how they were kind of applied to the programs that we identified.

So, the first on this list is the so-called conditional model and this is our name. It has been called different things by Tim, Ormond, and then Chris, in their paper a couple of years ago gave it a somewhat different name. But basically the idea is that you get -- there is a separate process through which cost is assessed, through which quality is assessed. And then there is this kind of joint determination of what is the cost for some given level of quality is typically how it is specified.

So, I think the key thinking about this is it is -- these aren't like independent signals but they are dependent signals and we are looking at the thing about efficiency by saying, by first kind of classifying quality

and then thinking about cost within that classification of quality.

So, I would say that is the essence of the conditional model. And this was used, I think for private insurers with tiering, how they created value tiers or whatever the distinction program that Blue Cross that Carole described and other payers are doing similar things. They typically use this approach.

So, an alternative approach that was, I think, more common, these kind of shared savings type models is, I mean it is related but they kind of idea is to set kind of a minimum threshold for either quality or cost, which is a so-called hurdle and then profile on the other dimension.

An example that came up before, I think Dennis mentioned, was the Medicare Shared Savings Program for ACOs and the kind of structures that there is a minimum quality standard and then there is this after, if an

ACO kind of doesn't hit that standard, then there is no shared savings, based on how they do on cost but if they do hit that hurdle, then there is the ability to share savings, based on how they do on cost.

So, these can be specified that kind of once you hit that hurdle, you are done and then quality kind of doesn't matter anymore. Or how savings is shared can be kind of weighted, based on how well you do on quality. That is kind of a variation on that.

And so we also saw kind of a cost hurdle model, where you start with some minimum threshold for how providers do on cost and then you start profiling quality above that. So, those are both variations that we saw.

So, another approach that was used fairly frequently was this unconditional model. And so the idea here is that you just have some quality signal, you have a cost signal and that they are weighted and

combined. But profiling isn't done kind of within specified levels of either of those dimensions. They are orthogonal and -- sorry, Joyce there.

CO-CHAIR DUBOW: I would like you to use plain English.

(Laughter.)

DR. RYAN: And so they are evaluated kind of independently and then combined through a waiting scheme. So, that is kind of what Hospital Value-based Purchasing uses now. And Alan invoked this before. But so in that program, so NQF actually endorsed the Medicare spending per beneficiary measure in that process. So, I forgetting the weighting but let's just say it is 30 percent of the hospital total performance score is based on how they do on cost. You know 30 percent is how they do on outcomes, 40 percent patient experience. That is not quite right but then those are combined to give a total score.

But that how cost is assessed isn't kind of dependent on any level of quality.

So the other model we have here is the regression model and I am not going to -- this is one that was proposed by Tim and Ormond and what I thought was a really excellent thoughtful paper. And it is trying to think about these, I think Larry said this before or it could have been someone else, maybe it was Tim, that is kind of wrong to think about quality and cost as being unrelated, that they are part of kind of joint production process. And if we don't think about that, the correlation and quality and cost within providers that we are kind of missing part of the signal. So, that is kind of the thinking behind the regression model is to kind of incorporate the within provider correlation between the quality and cost dimension and for how efficiency is profiled. So, there is a lot of technical details and I

refer you to the paper to sort through that. But it is a model that I think has some conceptual appeal but isn't currently in use by any sponsors.

And then let me -- yes, please, Jeff.

MS. WILBON: Microphone please.

DR. SILBER: Could you go into a little more detail on that? When you are talking about these being related, do you mean that the error structure is related between cost and quality and that is why you want to put them into a single model? Is that the reason?

Because that is a very different reason that suggesting that you do then what they did, which was report a conditional quality, based on cost. I mean there is two different concepts that are going on. One is a worry about the kind of metrics and the measurement and the other is just -- well, my concern is just what you get out of it is so

controversial.

So, would a patient care about what you get out of it? They might be glad that you got the error structure right but what you are giving them is something they might not want or understand.

So, I guess I want to separate those two concepts, what you get out of it and why they are -- you say it is nice that they are doing it with regression. Fine, they have somehow handled some of the error questions but they haven't -- but they are doing something that is very controversial in terms of saying to the public, well, okay, you do great, given that you don't spend very much, which isn't exactly what I think consumers want.

Am I right or wrong? Maybe I am misunderstanding it but I think that is the issue. So, I am not sure. We are praising them for a technical issue that is pretty minor, compared to what they are trying to do

with the model, which I think is concern.

DR. RYAN: Well, I think that is a great point. I hope that in the breakout session that we will have some time to kind of hash through the kind of implications of the technical specifications with respect to the use cases that we do consider.

So, and let me just continue on just at a high level to discuss another model that that same group proposed, which was to try to, and I think this gets us a little bit more towards this issue of value and that may or may not be a place where the group wants to go, but the idea with this cost effectiveness model is to try to put some valuation on the health benefits when considering efficiency.

And I think they are kind of the intuition here is that there could be a hospital that has say a much higher cost and slightly higher quality and, through an unconditional framework, could be considered to be inefficient. But if we consider if the

valuation of the benefits of that incrementally higher quality is very high, then they may, in fact, be producing health quite efficiently.

So that is what I think is an important consideration around, particularly around the units and how we think about the quality dimension. And you know if we are talking about an outcome that is mortality that we value very highly, small improvements of that on that measure could have -- societally could be very valuable. And I think this approach tries to get at this concept of the efficiency-producing health state, rather than of limiting it to the kind of more hard to interpret metric that we often kind of assign to say a generic composite quality of measure.

So, that is another approach that was taken in the literature. Again, it hasn't been put in use by any program sponsors.

So, let me just wrap up with the

final two. There is the DEA or stochastic frontier analysis. So, we heard from Peter, and then Steven as well, that this actually has been used in the VA to profile inefficiency. But it has been one, as Gary mentioned before that has had more kind of interest in traction on the kind of academic side.

And the basic idea here is that -- and Joe also relayed a very interesting anecdote which I hope he will share later -- we can define a frontier based on these inputs and outputs. And so what often has been done, and this was in the RAND report that Herb mentioned before, a lot of the kind of literature around measuring efficiency has defined output as kind of hospital days or some kind of standard kind of our physician used something like that. And then that would be the output.

And then we have something like physician labor or nurse labor or different

types of labor or capital are the inputs. And then you see which, based on that loose combination of inputs and outputs, there is a frontier. Then providers are profiled based on their kind of proximity to that frontier.

And I think that is some -- what is kind of nice about it is that efficiency is defined kind of over the range of all these combinations that we are not relying on. Kind of cut points around and thresholds to say that this group, these are efficient providers and these aren't efficient providers. So, I think there is some appeal there.

But I think, again, getting back to this idea of discrimination and interpretability, potentially one of the reasons why this hasn't gained a lot of traction in the field is that maybe kind of distance from the frontier isn't the kind of measure that is going to hold a lot of sway for patients or other people who are looking at these models.

And then let me just finish by talking about the side-by-side model. So, this is a model insofar as that there isn't a formal process of combining the quality and cost of dimensions to say this, to then kind of score providers in some way, based on the concept of efficiency.

So, this could be numerous measures that are shown together, numerous quality measures and kind of an episode cost. It could be a star system, which is being used more frequently. As Jeff mentioned, we didn't explicate all the ways that side-by-side models could be used but they are numerous.

But the kind of idea here is that there kind of isn't a judgment being made. It is just this is the information. This is cost. This is quality. And we are showing it together for assessment.

Yes, Peter?

DR. ALMENOFF: I guess for us, when we do the SFA or even DEA, quality isn't

part of that when we do it. After we have run the model, then we run quality against it to just make sure. High efficiency leads to bad quality would be, basically, a bad starter. So, the idea is to really show, either no relationship or a positive relationship.

But then once we have done that, then we actually will do side-to-side. So, we are doing, I guess, six and seven and it is also regression so far.

Because somebody already brought up this point. If I am a person wanting care and someone brought up I guess the Cadillac and what was the other?

DR. GOESCHEL: The Smart Car.

DR. ALMENOFF: It's not too smart.

So, a patient buying care might want a five-star for this and a five-star for that. But if you combine them altogether, there is no way of knowing what component they are interested in. One consumer might want a very good deal and doesn't really care about

the fact that it is not five, it is three. But to me, if you want to have a single score that is fine but you have to be able to break it out to know the differences because consumers want to know is the quality good and is the cost reasonable. And if we can't tell them that, then it is not really useful. Because as you know, you can have a 1-1 or a 5-1 or 1-5, the opposite direction, giving the same score but yet one is horrible efficiency, one is horrible quality. And I think somebody, a consumer really want -- and they will have a middle of the range score. And so here I was saying this is an average place but it could be horrible in one and really great in the other and they don't know that. So, I think it is important for them to at least understand those distinctions.

DR. RYAN: Right. And one thing I should note is that, and you just hinted at that, you mentioned it directly, Peter, is that there is kind of combinations of some of

these features for some of the models.

So, you could have a conditional model but you know the data are displayed, too. So, that is a good point. But some programs just kind of stop with that comparison.

Tim?

DR. LOWE: I don't see this so much as models as they are methods. So, I mean if we go back to our basic research class, I mean the research question and data drive the selection and the method.

So, what dependent variable are we using? Are we looking at cost? To me, I think we should be looking at variation as to the dependent variable, variation between providers, taking into consideration cost and quality. So, all things being equal are physicians, hospitals, whatever, providing the same care and kind of the unexplained variation. So we have explained variation and unexplained.

But my thing about the regression, I was going to ask you both which model would you or method are you arguing for or are you just kind of leaving it open?

And what is nice about the regression and I think, Jack, you probably would be pro the envelope. It is the most sophisticated, right? But the nice thing about regression is that there is a -- that you can do multi-level with it. So, we can look at a patient level and a provider level. Because there is a quality at the patient and there is a quality at the provider level.

And you had raised the idea, Jeff, about these error terms. And that is one way of parsing out the error because there is a mixing of those two, the organizational as opposed to the patient level. Because, obviously, there is an interaction effect. And that is one way of parsing that out, that error that you had referred to is by separating these into two level models. It is

much more sophisticated and I suppose it is probably, I think for the envelope analysis, as much as I like it, it is very difficult to explain it to people. Economists we can but to the average person, they are going to go, what.

But now that you have reviewed all these things, which are you recommending or where do you see, both of you see this going, in terms of actually doing the research? I am going to put you on the spot here.

DR. RYAN: Well, you know that is not -- I can just punt and say that is not our charge.

(Laughter.)

DR. TOMPKINS: I was just following orders.

DR. RYAN: So, you know we want to identify the universe and say what is there and, through this process, hopefully, get towards a consensus.

But I think that -- well, actually

I do want to stop there. Sorry, Tim.

DR. TOMPKINS: No, we are a team and I don't want to run over my teammate here. But just to say something that probably everybody knows and you started to allude to it and you did, actually, earlier, too.

To the extent that you are really going to focus on the technical properties of the modeling or derivation process, there are some elegances to be had, including DEA for its own sake, regression for its own sake, because of the error structures or the attempts to systematize the comparisons and so forth.

And so I have one of my colleagues at Brandeis is one of the world's foremost DEA advocates and proponents and I hear from him often enough to be convinced that that is an interesting way to look at it. Most of the world uses regression techniques or central tendency and, therefore, it is usually more tractable.

So from the point of view of measure development specification separating the errors and the variances into their levels, as you say and to be careful about not making inferences, for example, without respect to confidence intervals and likelihood et cetera is the territory of mistake and, therefore, more advanced statistical techniques can help guard against those kinds of things. So, we have to admire that there is a place for that.

Now, whether the place for that is in developing the original quality measure itself and the original resource use measure or cost measure itself, that could be the case, in which case we are still left with this conceptual problem of how do you bring the cost and the quality measures into some sort of nexus where simultaneously determinations are being made, inferences made across multi-dimensions. And I think that the focus of this meeting and the charge, as Andy

put it, is to contemplate the various ways in which, reaching back to Greg's pragmatism, the various ways in which measures, as they have been developed so far, anyway, can be legitimately or perhaps not so legitimately combined in a way that expresses validly what somebody is trying to say, namely, purporting to measure efficiency or value.

DR. SILBER: I agree with what you said and what Greg had said earlier. I just think that in the report, you could have, for the side-by-side model, you could have described methods of side-by-side that are meticulous in using the same patients that made the estimate for the cost as the quality. Or you could have looked at other side-by-sides that do things in aggregate and then hope that the model adjusts for it adequately. And you could do it by matching, which is another method that wasn't listed.

So, I just think that you have given us a lot of detail on one through six

and I think seven should be broken out into various approaches.

DR. PANTILAT: Yes, it's really great to see all these methods laid out here. From the conversation, just understanding that there are benefits to different ways of approaching this. But I worry a little bit that we might be confusing a little bit the method with sort of how it is ultimately presented to the end-user, whether that is the individual patient or purchaser or anyone else. And it seems to me it is possible to separate those two, that there might be a more sophisticated approach to analyzing quality and cost and even combining them in some way, and yet a very simple way to explain it.

At first I was thinking side-by-side makes a lot of sense to me to kind of keep things separate. But then as I was listening more, I thought I think there is a way in which we should use sort of the best method that produces the most accurate

assessment, and then think about how do we present it in a way that actually simplifies it so that someone can make use of it.

CO-CHAIR DUBOW: Dennis, did I see your card?

DR. SCANLON: So, just a couple of questions. One was about the time dimension and whether you thought about that or so much in the different approaches. So, you might think, for example, especially when you are concerned about both cost and quality, it is the cost trend that might matter. It is sort of how does a provider do kind of year to year and sort of what is the trend then in how -- or sort of how important is anyone, period, if you are measuring just at a single cross-section. So you can say the same thing in terms of quality measures as well.

If you think about use in decision-making, and this often comes up in the context of consumer choice, I get a report card based on a provider for HEDIS data that

was collected two years' ago or a year and a half ago. And so if I literally take that seriously and I make a choice, given variation in a provider, what is the likelihood that the decision that I make, which ultimately comes into play two to three years after the data that the choice was based on.

So, I was wondering if you might sort of talk about whether there was any consideration of time dimension in the things that you found or whether you think that there should be.

And related to that, one other -- well, not related to that. But one other question I guess related to the quality hurdle model. It seems to me that a number of public reporting models also take the distribution and create cut points. You know, top third, middle third, bottom third, or deciles or whatever the case may be. And often times, just make cut points to the extent I mean when we have collectively looked at the

readmissions rates, what we have found is that the cut points, actually, you could do that, but the difference between one decile and another actually is very small because they cluster. The distribution isn't as large as it is. It sort of clusters together.

So, I don't know. Would that type of thing come under the hurdle model? I think of the way this was presented in the paper, the hurdle was kind of a low-level threshold and then sort of above that, you start to sort, based on other things but I was wondering about that as well. So, really the time dimension and then this distribution cut point.

DR. RYAN: So, with respect to time I think there is two issues that occur to me. One is the one that Greg has been mentioning about synchronizing measure costs and quality measurement over the same interval and having a kind of joint measure. Probably the episode is the best way to think about it.

Where you have a defined episode, you have costs in there and you have quality in there and there is something very clean and neat about that.

We agree with that but we don't see that really happening in this space. I think one of the main reasons is that the quality measures that are out there, and this is a point that Carole made in one of our calls, there has been a lot of work specifying these and what is the right time period to look at all these things? When do you need this care? When do you need that care? And all those windows vary.

So, coming up with cost measures that kind of map on to all those, that fit those quality specifications is pretty -- that is a tough challenge. And I think so far the program sponsors have just said we are not going to be overly concerned. We think we are going to think about these signals as roughly reflective of what is happening in this

patient population. It doesn't need to map out exactly.

But with respect to like the time lags, you know wouldn't you say -- the two issues there are there is sometimes just a lag from program sponsors kind of getting, processing data, and getting it out there. And that is just detrimental to the whole process of profiling and public reporting.

But then the second issue is just for reliable signals, particularly for outcomes, you often need a longer time window for more cases to come in to be able to say with a degree of confidence that this is the point estimate. We feel confident that this is the level of quality from an end perspective.

So you know I think the issue, what we are talking about, is all those same issues that are out there in profiling costs and quality. They are still out there. These are, I want to say, kind of generic issues

with public reporting. And I don't think we are any, frankly, closer, to addressing them but I do agree from the perspective, particularly from consumer choice that these kind of lags in information being used for profiling inhibits informed choice from patients.

CO-CHAIR DUBOW: I just want to call attention to the time. So, I think that the tents that are up, we can talk about what we need to get to the breakout groups.

So, I have -- did you answer both of Dennis' questions by the way? Dennis, did we? We talked about time and there was the cut points.

DR. SCANLON: Yes, I think it was good enough.

CO-CHAIR DUBOW: Okay, Matthew.

DR. ROUSCULP: So, I guess going back a couple of elements. So, first of all, from a pragmatic perspective, it sounds as if we want to be able to capture variables that

are relatively easy. If we look at NQF measures, these are things that are kind of straightforward and direct and it is usually at the organization level of provider level.

But the question is when you start going into the regression model and some of these elements, especially when you are talking about quality and patient level quality but there is some of those patient-level variables that are going to be very important. That is kind of like ultimately where we would want to go but we have to be pragmatists here.

So, from your perspective, I was just wondering understanding some of the limitations as good as any of these models or the analyses that are going to go into these models, whether there are some of these things that we can say look, it is a good approach but it is not good enough. Therefore, the results are going to be shown is more about sensitivity analyses. And it is even with the

conditional model of others. We are going to have to present the data with a very wide range that is occurring. We need to be considerate of that, especially if we start talking about ratios. So, now I won't have the cost but I will have quality. We kind of mesh it. That kind of gets us really into some ugly problems as far as what are our results.

And that then gets back to, do we need to be considering some of those things to be able to say look, when we start doing these analyses, there isn't a certain number. There isn't going to be something clean that comes out of it and is that okay with us.

And the second, I guess, is around a lot of the discussions and more about a static approach. Right? It says what is the value as of whatever is being measured. Is there a dynamic nature, to expand upon what Dennis was talking about, whereas, timing. Is there change over time that we are actually

seeking and is that of value? So, you may not be one of the top performers but by moving upwards and becoming a mid-level performer in quality or being the high cost and going down to not such a high cost, does that actually have value in that because you are moving in the correct direction?

DR. RYAN: I'll take this. Okay, those are both excellent points.

So, the first one about how should we manage uncertainty in the estimates, you know I think this really is the way that measurement has been going is that a lot of measures are trying to do this at the measure level. So you see how the kind of measures that CMS has -- that have been endorsed here for mortality and readmissions. What they are trying to do is trying to take the noise out to the extent possible through this shrinkage approach, which Jeff and others are thrilled about.

(Laughter.)

DR. RYAN: And so at that measure level, they are trying to manage the uncertainty. And so but that is done to a varying extent across different quality measures and typically process measures. I have never seen any kind of shrinkage applied there. There is often just kind of cut points.

But the issue as to how do we say whether the output of kind of the quality and cost combination is sufficiently reliable, I think that gets into how NQF is going to -- it gets into the kind of guidance that we are going to be giving them for how should we -- what should the kind of recommendations be with respect to how should reliability of an efficiency measures -- should we apply those same kind of criteria that we apply to individual measures with how those two things are combined in the output of that model.

And I, personally, think yes, we should. And a lot of that same testing that

we could do with reliability and validity is come up, I think, Jack and the other people have mentioned that the different models could give you different rankings and a way to show a model could be kind of a validity check could be if different models are giving you kind of a similar rankings of providers.

So anyway, I think there is kind of standard methods that we could use to apply to the combination of these measures that would kind of speak to the issue of reliability.

On that second point about improvement, so I know Hospital Value-based Purchasing, I know the program well. So, that is one where there is incentives for both attainment and improvement. It kind of gets rolled up into a total performance score for on the measure level, the maximum points for improvement or attainment are assigned and then hospitals get that. So, I recall there being one other program that incentivized

improvement but from most of what we have seen, it is really we are talking about levels that are being kind of profiled incentivized. So, it is kind of like how we first thought about pay for performance just profiles based on levels and leave it at that.

So, I think, again, I am not -- I can't speak to the entire universe of burdens but I think the emphasis has really not been on improvement to date. It has really been on levels of performance.

CO-CHAIR DUBOW: Although, I think in some cases that improvement, attainment business is statutory. Congress has looked to that as a mechanism for I think appeasing some constituents to give credit for improvement, as opposed to having to hit a mark.

So, I urge you, please, let's be concise so that we don't go too far over. And we still haven't heard from Chris, which I want to do after we run through our round.

Jack?

DR. NEEDLEMAN: The point that I don't think has been made is that there is an analytic relationship among all these methods, including the regression-based methods. And the way to think about that is, to me, the way I think about it, is I think about the scatter plot of however I have measured quality and however I have measured cost. And each of those have an uncertainty about them because of the uncertainty in my data and a lot of the arguments we have over the measures and a lot of Jeff's concerns about shrinkage or how we position those points, given those uncertainties in the data.

So, we have got the issue of have we got the scatter plot right or are the points moving around when we measure things differently. And that is one set of issues that is across all of these measures. But if you think about that scatter plot, what we are seeing with the conditional, the side-by-side, basically we are drawing lines in that scatter

plot and saying this quadrant is good and this quadrant is bad. And we are potentially paying people who are doing that. And that works.

And an awful lot of our analysis is driven by the fact that whenever we do these scatter plots, the space is reasonably full of points. So, what conditional model, what the side-by-side do, is they basically divide things up nonparametrically and say where on this graph of scatter points is your place. And is that a good place to be or a bad place to be?

What the regression line does is it puts a line through the middle of that data cloud and basically says who are positive deviants. And what the stochastic frontier analysis as the data envelopment people do is they put a line around the outside of that cloud and they say how close are you to that line of possibility and, preferably, how close are you to that line of possibility at low

cost.

Now, with the clouds completely filled, that DEA or frontier line is going to swing over the whole top of the thing. And we have got low cost providers that are delivering high quality and high-cost providers that are delivering high quality. And that is what the DEA will tell us. That is what the conditional model will tell us. That is what the scatter plot will tell us. And an awful lot of our decision-making around this, built around the assumption we are going to have a pretty scatter plot where we have more troubles and more difficulties is when that upper left quadrant, low cost high quality is basically empty. And the only way you get higher quality is to pay more.

And it is in those models that the regression models, the data envelopment, the stochastic frontier tell us something useful about the nature of those trade-offs in ways that simply drawing the lines and say I am

going to circle this group doesn't. So, we need to think about the models. We also need to think about what the distribution of quality and cost are and how uniformly distributed along both dimensions places are, because that affects the way we use it. It affects the way we interpret it.

It is like if you are doing your -- it open enrollment and all the five-star health plans are $400 a month more than all the four-star health plans. It is a different decision than if you have got a cheap five-star and an expensive five-star. And that is what we are dealing with. And each of those scatter on those two dimensions differently. That is what we are dealing with here but there is a relationship among all of these approaches, in terms of they are looking at the same data and they are trying to give us different metrics for interpreting.

MR. ROMM: So, I agree very much with all of those points and that is the

pathway that I was actually heading down. So, I will not say any of that, other than to say absolutely.

I think that on top of that, there is another interesting consideration as we start to think about use cases, especially. Larry and I were talking about this a bit at the break. All of these models, except for maybe the side-by-side were developed for the purpose in healthcare effectively of payment. They are not being used for anything else at this point, really. There is a little bit of public reporting in small pockets but not much.

And so I think that then when you recognize that on top of that, fundamentally, all of that payment is in negotiation in almost every setting, except for Medicare. Then, it become somewhat of a useless exercise, I have to say. And so now, when we are talking about a commercial market, I think that we have to start to set some expectations

about where we want efficiency to go, which means making some very hard decisions about directionality here. Because otherwise, it is a fairly useless exercise and it becomes a negotiated base, negotiated performance targets and there is not much there.

CO-CHAIR FLAMM: Just one quick response, though. I think, especially for the designation programs and those that are listed in there, they are not actually for payment. They are for informing benefit designs and member choice. So, I think that is a little bit of a different lever, than payment per se.

CO-CHAIR DUBOW: But it is payment because you have set a value to it.

MR. ROMM: Exactly.

CO-CHAIR DUBOW: It has an effect of payment.

MR. ROMM: Agreed.

CO-CHAIR FLAMM: Okay, sorry.

CO-CHAIR DUBOW: Okay, Peter and Alan, and then Chris.

Jeff, is it really pertinent? All right, quickly.

(Laughter.)

CO-CHAIR DUBOW: Is it apropos to this point, Jeff? Are you commenting on this observation?

DR. SILBER: I was going to be brief. But I was just going to say I think we are missing the forest from the trees. And that if you go to a simple example, like look at observed over expected, when you look at that metric in Health Services 101, you don't put hospital characteristics into the expected part of that ratio. If you put, say nurse to bed ratio in that expected part, you would say oh, this hospital is great, given that they have a terrible nurse to bed ratio but the death rate is terrible.

I think we have to keep remembering that that is the problem. And as we talk about efficiency in method one through six, we have that problem. And I just feel

like the conversation is missing the point and that we are going to get into trouble if we use one through six and give that to the public. It is like saying, they were a great hospital, given they had terrible nursing.

CO-CHAIR DUBOW: Peter.

DR. ALMENOFF: One quick comment and then a question.

Somebody had asked which model is probably the best model to use. And actually reading the study, the white paper you wrote, I think that something that would be extremely beneficial to everybody is actually just list all the possible models that exist, what their pros and cons are, what their use and strengths are. That would, actually, be a lot more beneficial than us trying to figure out which model is the right model. Because I don't think any of them -- I think all of them work in the right situation and in the right hands.

So, just the fact that we know

some of these models clearly are not going to work in certain scenarios, some of them were built for the payers. Some of them were built for the providers. Some of them were built for big health systems that deliver money to hospitals. So, they all have pros and cons. But I think if we could just outline that, that would be much more useful to the public or anybody trying to build these measures that they don't try to give us an SFA model and an individual provider reimbursement system because it is probably not smart to do it that way.

So, I think that would be beneficial because I have not actually seen it -- I know about most of these models but I have never seen them altogether and it was actually a very nice summary for me.

The question I had was one of the things we are struggling with and maybe you know from some of the other issues is it is very hard for us to look at this over time

because we sort of have risk adjustments. We adjust across the country. As I mentioned, we look at things like snowfall. People complain we have more snow than other places in the country so we are less efficient. By the way, that is not true. We check it. We look at the snowfall in every place in the U.S. and there is no relationship.

But when we do that, every year the models does change a little. And the purist in our group say we can't really say that if a place was eight percent efficient last year and is six percent efficient this year, I can't actually say they are better or worse.

So, I think a lot of these models are going to have those similar issues. And so if we are looking at a fixed period in time and we are deciding a lot, then a provider or health plan gets better, I'm not sure how to reconcile all of that.

DR. RYAN: Right. If all of the

comparisons are relative, then how do you say every year that there is a real change?

DR. ALMENOFF: So did you come across anybody who was able to address any of that? Because we haven't been able to figure it out.

DR. RYAN: The example you just gave, I mean it sounds like if there was just a frontier that was joined over both periods, then you wouldn't have a shifting frontier in both cases. I think you could assess whether there was a kind of difference in that distance in the two periods. But no, I can't think of -- nothing comes to mind that explicitly addresses that.

DR. ALMENOFF: Then the last comment was, people mention a lag of quality. What we have is a lag of the financials. So, we have a harder time getting accurate up-to-date financials. And we only can build this thing once a year because the data, the financial systems are not stable. And I don't

know if that is the same in the private sector but I have a feeling that they are the same.

So, we are having the opposite issue of we have very accurate quality data, pretty current, but we lag the efficiency side.

CO-CHAIR DUBOW: Okay, Alan, very quickly, I hope.

DR. SPEIR: As we move into our breakout sessions and we are going to be discussing these different models and the components of them, there is a fundamental question I still have that it has been touched on particularly by Jack and Jeff but I still don't get it. And that is, we can understand outcomes. We can understand and agree upon how those transcend variation.

How do you normalize cost? So, how am I to understand that a model that works in one locale and if I build, in your words, an efficient frontier that would be established and then the provider efficiency

will be based upon the variation from that frontier. But as I mentioned earlier this morning, given the variation in cost and how that is defined by the different payers in the different sectors, how does that relate and how can we get a model that will then be able to be understood globally, absent the normalization of these costs and how that relates to these models because I don't get it.

DR. RYAN: It is, I think, something that we just alluded to here. But we have heard a number of comments today about normalization. And we heard from the Massachusetts experience, it is an interesting question how much do you want to normalize, how much do you want to standardize? I mean, do you want to strip away price differences that are done on the basis of negotiating leverage? Probably not. Do you want to adjust for regional wage differences? You probably do.

So, I think --

DR. MAC LEAN: But why -- I'm sorry. So, let's talk about say I am an employer and I have got this group of employees and I have to pay for their services. And if it is cheaper for me to send my employee to a hospital in Ohio than it is for them all things included, airfare and hotels and the like, if it is cheaper for them to go there, then why not consider that, than to have it done in Los Angeles?

PARTICIPANT: Because I just don't want to travel. Patients don't want to travel.

DR. MAC LEAN: Well, actually, we have patients who do travel.

CO-CHAIR DUBOW: We are not going to resolve that one. That is a very complicated question.

(Laughter.)

CO-CHAIR DUBOW: I just want to give Chris the chance to add any commentary or

make any other observations before we hear from Ashlie with our directions for the breakouts.

DR. TOMPKINS: Sure. Well, given the hour, what I will do is I will try to do a mini segue, successfully or not. I don't know.

One of the things I have been envisioning is the pinwheel, merely to make a point here, which is that there is a heart to the issue that we are having to address as a committee and as authors. And distinguishing that heart from all of the strands that come off of it; i.e., the pinwheel.

And I think that even from our earliest conference call and comments, we have all agreed that you can't have the heart alone without the pinwheel. That is, even if it is just a reference for sophisticated readers, we have to say that there are other important issues that connect here, even though the purpose of this paper or the purpose of what

comes out from NQF isn't necessarily handling them with sufficient thoroughness and so forth to be topics in and of themselves. But they are reminders, signals to the reader that note that this is an important related concept. So, as you think about the heart of this, go elsewhere, too, and consider what else is out there that impinges on the heart of the pinwheel.

The second comment, which is quite different, maybe just tell a quick story. It begins with -- maybe some people remember this and maybe some people prefer to forget it. The word is RHQDAPU. Does that ring a bell? It probably does because if you have heard it once, you have probably heard it a thousand times. And if you haven't heard it at all, be glad. It is one of the government acronyms, Reporting Quality Data for Annual Payment Update. Is that rate? It was the hospital reporting requirement, which still exists to produce the aspirin on arrival measures, the

beta blocker measures and all the other skip measures and so forth that is now, fortunately, under IQR, instead of RHQDAPU.

The Hospital Value-based Purchasing is one of the models here. And I actually kind of led the development of that. And I had some help from around the table. Andy was with me at the time and Gary Young, at the end of the table, he was part of my team. And we started that when there was RHQDAPU. That is, there was a public reporting program. But the charge was to come up with a Value-Based Purchasing Program. And the question is whether you pull it over from now what is called IQR versus what is different.

All right, the reason I am saying this partly is segue is it would have been nice to have this paper. It would have been nice to have this committee's considerations and outputs back then. Because, for example, when you started out with just the public

reporting, you can have any number of stand-alone measures and that is what they did. You can go to the -- say here is one measure. Here is another one, for what it is worth. How do you like this one? Here's another one and here's another one. And it is the ultimate side-by-side, at least on the quality part because there was no attempt to comment on which is more important and which one is more reliable and which is more discriminating. It is just like pull up the hospitals in your area and you can see how well they deal in this sort of thing.

And for public reporting, it is actually okay, for example, if everybody does well in that measure. It might be reassuring. It might be informative. It might, in a sense, be useful. But, going from public reporting, which is just this full disclosure sort of transparency model to value-based purchasing, what you have to do is to say we are actually going to start to cause -- we

need some summaries here. We need to integrate these measures and we need to eventually come up with a singular measure, which we call the total quality measure Andy referred to with the attainment and the improvement, and the point system and so forth like that. It was an attempt to take disparate measures that otherwise did and could stand alone and turn them into a singular message.

And without elaborating this sort of too much with the domains and so forth and so on, one of the things that we discovered was that, as I started to say, some quality measures, if everybody does well, is fine. But if you are trying to discriminate, you can't really use the -- you can't necessarily use the same set of measures because if some -- just because you can rank order providers on a measure, if it is a highly clustered distribution, the differences aren't meaningful. So, what you end up doing is

splitting hairs. Yes, somebody scored 93.0005 versus 93.0004. So, therefore, one is better than the other? No, I don't think so.

So, the implications if you are trying to start to distill for payment purposes, or particularly a use case, you have to decide the quality of the measures and what they are really saying and how much it really brings and new useful information versus whether it is redundant or whether it is just bringing noise and it is confusing the story.

But then came along the ACA and it insists that the CMS develop efficiency measures. And now the Agency goes what is that. And that is where this paper could have been useful in part -- well the paper yet to be written of course -- where the paper that we are envisioning could have been useful because one of the implicit questions there was to say okay, you have got these domains already in quality. You have a process domain. You have an outcome domain. You have

a patient experience domain. And they are weighted, the preference weighting for value and so on.

Well, let's just make another one. Let's call it the efficiency domain. It doesn't matter what is in it. It is just that this is going to be where the resource use and the dollars are in. And so candidates started to be well, one measure that could go in there would be the emergency room throughput time and maybe blood units per surgery. In other words, resources that are actually used in the ingredients of production.

Okay, what turned out to be the case because ACA also required the hospital spending per beneficiary measure to be calculated, which is probably familiar to some of the heads that are nodding and others as well. And we started to work on that. We worked on the first version of that. But that became the candidate for what had become the sufficiency.

But the question there in the efficiency became, particularly as it related to hospital spending per beneficiary was, this is now the 30-day cost measure, it is the DRG plus the 30-day post-acute. The question became -- now this goes back to what I said in the morning about being tethered or better or for worse, so the previous definitions. If the definition that NQF, AQA, and so forth were saying what efficiency was, cost in relation to a specified level of quality.

Then the question is, if you just had cost as its own separate domain, standing alone in what is called here the unconditional model, is that good enough? Or, to be compliant with ACA's requirement for an efficiency measure and to also be faithful to what the industry is saying is what an efficiency measure is, do you have to link the two?

In other words, so here is a practical question. If you are going to

compare the resource use or the cost of a given set of hospitals, should it be on average across all hospitals or should the comparisons be limited to hospitals that are said to provide the same level of quality, which comports with the definition? And if you do that, then implicitly, the cost distributions could be different, if you said the highest quality hospitals will only be compared to highest quality hospitals.

And if it turns out that it is cheaper to produce lower quality hospital products, that doesn't matter for this comparison because we want to know which hospitals among those that produce the highest quality are the most or the least efficient.

And I think it is sort of in this realm of whether or not you need to consider, and that is the practical question, does the benchmark for cost, and the distribution, the implicit distribution for comparisons, be made in light of a specified level of quality or is

it okay to say it is a separate unconditional domain?

And so, therefore, the segue part of my comment is that if we are going into the breakout sessions, it is a question about what are you trying to accomplish here. And then it is not so much do I like regression or do I like DEA. The point is from the perspective of this stakeholder, the person who reads this report or the person who is actually going to try to use these measures or make decisions along these lines, if you are trying to do it for public reporting, take that perspective and decide what is the question. What are the most important questions to tackle, the priorities, what is important, what is less important, if you are thinking about public reporting. And then another group might be thinking about network management and so forth.

And I think that what we want to do is to have everybody bring their best ideas

into whatever group you are in, make sure you heard but use it within the constructs of that use case and see whether or not, at least to generalizable conclusions, they could spill over to other use cases or maybe you identify things that are really particular and special for use case. And so, therefore, that has implications about how we might, as it were, stratify our recommendations.

CO-CHAIR DUBOW: Thanks very much, Chris. Ashlie?

MS. WILBON: Yes, thanks. Chris actually hit on several things in his comments, hopefully that will actually be a very good segue into our breakout discussion.

So, you probably saw us huddling before lunch started. And we have really been kind of digesting everything that has been said today. There has been some really great discussion and we were trying to figure out the best way to integrate some of the comments that you have and also kind of give you the

best direction that we can, so that are you are able to continue to give the rich input that you have in the context of what we are trying to get out of here and make sure that you have clear direction as well.

So, what we landed with is we are going to continue to have you break out into groups based on use case, and we will share the assignments shortly, into four groups. But we have kind of tweaked the questions that were in the discussion guide a little bit. We are going to have -- one of the suggestions that we are going to go with and see how this works because a lot of this has been somewhat of an organic discussion and obviously, we came with a plan but sometimes things change and we want to be adaptable and flexible and go with that so that it works for the rest of the discussion.

So, rather than you have guys focus specifically on the models, per se, and how the models fit a particular use case. We

are going to have a little bit broader discussion around the use case and what types of considerations you would have for measuring efficiency within that use case. And if through that discussion you come up with, as Chris described, for you to do this particular use case that you need to have measures or an approach that does A, B, C, and D, that perhaps you end up describing some parts or pieces or a full model that has already been identified in the environmental scans, we are trying to break you free of having to stick to specifically a model but have you think broadly about kind of what you would need to be thinking about for that particular use case.

So, hopefully, that resonates with people and will help kind of facilitate the discussion a little bit more so you don't feel constrained. And there will be a staff facilitator in each of the groups. So, we will try to answer questions. We are going to

have Chris and Andy, as well as the co-chairs rotating between the groups. So, to the extent that you have questions, hopefully they will get to every one of the groups.

We are going to have the groups kind of move to different sides of the room so it is easy logistically for people to rotate, for our chairs and authors. And we may end up shifting a little bit. And it is not particularly ideal because, obviously, you are going to hear discussions of other groups but we will do our best to work with it. If it gets too noisy, we do have an alternative. We can move to a space somewhere else upstairs. But we are going to do our best with the space and, hopefully, we can work that out.

Again, we will identify a spokesperson to report out. And I think, given the time, we are going to -- I think the agenda had us only going to 3:15 for the breakout sessions but I really feel like we are going to need more than 45 minutes at this

point to discuss the different use cases.

So, why don't we go to -- would people be okay with going to let's see, 3:30 at least 3:30 or 3:45? We will do a pulse check around 3:30 to see how groups are going and if you guys need a break if you can go another 15 minutes, we will keep going and we will kind of go from there. Let's say 3:30 for now and we will pulse check.

I do just want to go through the question that we are posing for each of the groups. Again, we are trying not to constrain you too much but we want you to have some structure. You have got to have an idea of what we are looking to discuss. Many of these things are some of the kind of topics that Chris touched on in his last comment.

So, for each use case, we are looking for you to talk about the considerations for measuring efficiency for the use cases, which includes the audience, interpretability, extent of the scientific

rigor that may be required and then the four use cases that we discussed earlier, quality improvement, public reporting, network design, or tiering and then pay for performance.

And then again, how might the intended perspective of the intended audience impact the selection of the model for a particular use or uses. And then Erin, could I have you go to the next case? There should be another. There we go.

And then what, again, principles should be considered when selecting individual measures for cost and quality into the use case. So, again, how are you deciding which measures get put into, for that particular use case, how are you deciding which measures you are picking to measure efficiency?

So, we have some examples here of some things you might want to consider. But again, we are encouraging discussion to the extent that you have ideas about this, measure type, whether or not there is a performance

gap. Is there room for improvement? What the focus of the efficiency measure is to be. And then stakeholder perspective, which we have a lot of discussion about already as well.

Then, I think, there is one more set of questions around kind of the need or lack of need for technical or conceptual alignment of the various measures that are in the actual efficiency approach. So, do they need to be aligned in terms of the developer having developed both a cost measure and the quality measures at the same time and sharing the same denominator, measure population, risk adjustment and so forth.

So, again, this is not an exhaustive list. This is just, again, to give you a framework of what to be thinking about. We encourage a rich discussion to the extent that you can, based on your use case that you have been assigned to.

Does anyone have questions about where we are going with this?

CO-CHAIR DUBOW: I think we're going to just try to get the question to each group.

MS. WILBON: Yes, we will print them out for you. And we changed a few of the questions but for the most part, they are in the discussion guide. So, if you have your discussion guide, it will give you a pretty good idea of what we are going to be talk about and your staff lead will help facilitate that as well. So, but we will also make sure that you have an updated list of those questions.

CO-CHAIR DUBOW: I just have a very quick question. Does anybody have an absolute hard stop at five o'clock?

Matt? Okay. Did I see somebody else? That's two. Okay, thank you.

MS. WILBON: Vy, can you share the breakout group assignments, please? Thanks.

So, Group A we will put in the corner behind us to the left. That

facilitator will actually be Lindsey -- not Lindsey. It says Lindsey but it will be Erin. Group A up here in the corner.

Why don't we put Group B up here in the right-hand corner with Rob, another staff person here? He's over here.

Group C, we will put in the back to the right. That is with Taroon. And then Group D is pay for performance, that is me, will be back by the food and the coffee.

(Laughter.)

MS. WILBON: So, we will give everybody a couple of minutes to get to their groups.

(Whereupon, the foregoing meeting went off the record at 2:29 p.m. and went back on the record at 4:02 p.m.)

MS. WILBON: If everyone could start making their way back to the table, we would like to go ahead and get started with the first report out, which was, I think we are going to go in order, starting with Group

A.

CO-CHAIR DUBOW: So, can we bring everybody back together so we can -- we don't have enough of a quorum, I don't think.

So, this part of the afternoon is devoted to the report outs from the first breakout sessions. Is that right, Ashlie? And do we have -- if we go in order, is the quality improvement group here and is your spokesperson here? Well, I actually know you are.

DR. WOZNIAK: Should I stand up?

CO-CHAIR DUBOW: You can do anything you like, just so long as we can hear you.

MS. WILBON: Hold on one second. Erin, did you guys have slides? We don't have slides for our group, by the way.

CO-CHAIR DUBOW: But you know we are going to need to hear you.

MS. WILBON: Hold on a second. We are going to give you a microphone so it's

recorded. Thank you.

DR. WOZNIAK: Well, I don't know if anyone read that New York Times article a couple of years ago about how in the Army they don't do anything without PowerPoint, the whole PowerPoint culture. So, hopefully, we can do something with that PowerPoint.

Our group had a really interesting discussion. The first thing we discussed was actually what is our question. Because we were not certain if the question was what measures are useful to determine the efficiency of quality improvement efforts or what measures will be useful for assessing performance improvement efforts to improve efficiency, efforts to improve efficiency. And we decided to focus on the latter. We thought what we were looking for was measures and models that would be used as the measures when providers, organizations, were doing projects to improve their efficiency.

So, the first observation, an

observation we had is that there is a difference between measures for our perspective and the three others. The three other perspectives are largely external, either patient or payer perspectives on healthcare delivery and total cost as price can work for those perspectives. But for the perspective that we were talking about, it really depended, you really wanted to get the economic efficiency of the cost of production of services by the provider and match that with the quality of those services.

There was a thought that the metrics are going to need to be actionable. And the perspective is very important. And this perspective would be the healthcare delivery system or provider. So, we should probably give some examples. But if the example was the cost to produce a cabbage -- so stepping back, we thought the measurement perspective was largely for an organization or provider that was ranked in one of these other

systems as an efficient system, as an example, and wanted to improve their rating on the tiering scale.

MS. WILBON: I actually think your microphone went out. Try it again, maybe. We would like to get it on --

DR. WOZNIAK: So, the metrics need to be actionable. Many of the measures that have been developed are not going to be actionable at the provider or the institution level. So, the example of the -- some of these examples, a tiering program that tiers hospitals in three levels of quality, that is generally not going to give the hospital actionable information on how to improve their efficiency.

And so, measures in this domain are going to need to align cost and quality measures around episodes or conditions that are actionable. And data is a challenge in this space. Most American healthcare providers don't have accurate cost accounting

information and so that is another area that is a challenge.

We thought that outcomes measures would be best. So, the true production cost. And if you read about Intermountain Health and Brent James and all of the work they have done, they have a very accurate cost accounting system and so they can look at their true costs. But in many systems, we don't have that. And so process measures would be important, things like average length of stay for admissions, nursing ratios, and other process measures of cost, as long as those are not being done or improved at the expense of overall cost.

Because we had an example that was if you wanted to improve the internal efficiency of care for patients with pneumonia and you had long length of stay, you might hire a hospitalist service. But the cost of hiring a hospitalist service, unless you are accounting for that, you might be able to

improve the process of length of stay but you need to measure cost to get the whole part of that.

Having compatible benchmarks was a key issue. And so, there are currently organizations, ActionOI, UHC Premier that do these sorts of measures for improvement on efficiency internally. And the hospitals in particular get their data because they want comparable benchmarks. But the challenge is that since our cost accounting systems aren't accurate, often even these metrics are not really comparable across institutions. So, the way in which nursing hours are accounted or in which other pieces that go into these are accounted may not be comparable across institutions.

And so for the models selection, I think that was the next one, we thought the potential model that would work would be the side-by-side model, where an institution would look at -- and this I how many institutions do

this now. They have a dashboard of some type that has internal cost measures, either process or actual cost measures and also has quality measures. And the people who are doing the performance improvement look at both of them.

And that the quality hurdle framework would probably be how most places would approach this. They wouldn't want to decrease quality in order to -- to decrease quality less than they are decreasing costs. So, they would want to hold quality constant and usually at some level that they had set as a goal.

You could imagine using a method like DEA if you had a network of providers and could compare lots of different providers, different hospitals or entities to look at across the production frontier how they would compare. But as an individual institution, it would be hard to do that.

So, I think that is the summary

from our group but I want to let other people from the group speak up because we had a diverse and interesting discussion.

I guess one last point, our last point of debate was is this the space that NQF should be. Should NQF be in the space of reviewing and endorsing measures which are essentially for internal improvement? And NQF used to do that a long time ago but now all measures have to be outward facing also. And we thought this probably is not actually an area for NQF to be active.

And so we thought it was important for the product of this workgroup, the paper to discuss the limitations of current efficiency measures in providing actionable information to improve the actual efficiency of care.

CO-CHAIR DUBOW: So, do any of the members of the breakout group have anything to add?

So, let's have a -- are there any

questions from the rest of the workgroup?

I think I would like to ask NQF staff about the very last point that was made with respect to what NQF's interest is in having your measures alone.

MS. WILBON: So, I'll start. So, we are actually going to have a more in-depth discussion around the endorsement process tomorrow, which I think is where that discussion probably goes. But he is correct. In generally, we have our policy is that we endorse measures for performance, improvement and accountability purposes, and not only internal QI. I mean, there are measures that can be used for accountability purposes and performance improvement that are also used for internal QI but not just approaches for internal QI.

I will say in the concepts of the paper, though, while endorsement may be off the table for those particular applications of efficiencies, approaches or models, that type

of thinking around that context may be useful to other entities that would pick this up and say maybe NQF is not going to endorse our efficiency approach but maybe the approach that we use at our institutions, we can take some of these principles and apply it.

So, I think, obviously, the work of the group is still very valuable. But in the context of NQF endorsement, perhaps --

CO-CHAIR DUBOW: Although there has been -- I am just blocking on the term we used but there has been some discussion at NQF about looking at a continuum of measure types and being very clear about purpose of the measure and slotting them. I can't remember what we call that.

So, there has been some discussion about that but I think the point is well taken and that we should think about that.

If there is no further discussion or questions about this, we can move onto the second group, which was public reporting. Who

is the spokesperson for that group?

DR. SCANLON: Okay, so I am.

CO-CHAIR DUBOW: Dennis.

DR. SCANLON: I was nominated and we sort of passed the baton a few times but I think that all of us had enough uncertainty to some extent that I am going to pause at a few occasions and let others on the group sort of chime in so I don't screw it up.

So, public reporting, obviously, is a very common use of measures reporting and we wanted to sort of take a look at this in the context of efficiency. We really sort of answered these questions. Why would we publicly report? We really found ourselves after the fact kind of coming back to that question and saying what is the value of doing this. I'm not sure that that was necessarily on the list. What do we hope to get out of it, how to involve stakeholders, what type of efficiency could be reported and what is needed for the future if NQF and others are

going to go in this direction.

And so the question why would we publicly report efficiency, the credible threat of consumer use in many respects when it comes to provider quality or patient experience, I think there are some who sort of wonder whether consumers, patients look at this information, whether they use it, whether they make decisions based on it but there is the view that there still is value in the market place for purposes of improvement or for providers to kind of benchmark relative to others and consider where they stand and make changes as a result of it.

You know, it is, in specific cases, things like labs, imaging, there might be opportunities where people would use this information. We spent some time talking about kind of the choice situation from a consumer perspective in considering specialist referrals or acute and emergent conditions, versus conditions where you have more time to

make a decision.

There is a lot of factors that would potentially prevent consumer use of this information or patient use. But there are some circumstances where it might be more likely than others. There is certainly the provider reputation reason and the notion that when things are reported, they become public. They get disseminated through the media, through the social media and reputation can be affected or impacted by virtue of how one looks in these types of reports. And as a result, that may motivate providers as well. And there may also be some research value as well.

So, anything else from our group's perspective on this question?

DR. WONG: I think that is a fair characterization. I think our group really started from the point we had some doubts whether or not consumers would really look at efficiency measures. So, we came across the

notion or recognition that this is a train that is kind of moving forward. To the extent that we want to contribute to the conversation, what are some of the reasons why we believe folks might looking at that. And so the credible threat sort of aspect came up and the provider reputation sort of came up. And that is how our conversation progressed.

CO-CHAIR DUBOW: Dennis, if I could just take my chair hat off for a minute. You know, again, if we think about linking cost and quality measures, as opposed to thinking about efficiency measures, then I think there is a clear consumer interest. So, I think again this becomes definitional or focus or however we characterize this. But I would not want to take consumer interest out of this public reporting space without some acknowledgment that cost and quality together represent a very important area for consumers.

DR. SCANLON: Yes, I think that gets back to sort of definitional sort of what

it is being sold at and a whole variety of things that we have talked about today. I mean is it a composite with a name like efficiency. Is it side-by-side. But I guess maybe you are making a case for or possibly making a case for the side-by-side component where quality is very clear. And then there maybe this other piece, which may or may not be as actionable.

CO-CHAIR DUBOW: And I think we can't lose sight either of the sheer value of putting sunlight on performance and the well-known responses from the clinical community to doing that. It has driven improvement, even if consumers don't use the data themselves.

DR. GOESCHEL: I think, Dennis, if I could, just to speak to that, I think we get at some of that as we get closer to -- even as we went around the circle, I think we got some of that. So, I agree with you.

DR. SCANLON: Right. So thinks like this, getting consumers to ask questions,

have greater awareness of variation, maybe, as we said, select efficient providers. So that is kind of the area I think that creates the most uncertainty in our mind.

You know we even talked about things like from a public policy or sort of public goods perspective the notion that -- just the fact that there is variation or there may be variation in efficiency in expenditures due to that from a publicly funded healthcare program perspective as a series of tax payers, consumers might be interested because their money is potentially being spent on less than optimal or less than efficient providers. That might require a packaging of information that is a little different. In that case, it is the message. There is variation and there is inefficiency and somebody is paying for it, as opposed to sort of information in a traditional kind of report where you might be using it to make a choice.

And then certainly to have prices

converge to rational pricing is a vehicle to sort of have things converge. We talked earlier about the example in reference pricing in California.

For patients, the information would be better if it was easily understandable, had face validity, was meaningful, displayed in a visually understandable way, so less technically complex, perhaps, or at least displayed in a way that the technical complexity of the computation would not overwhelm the presentation and from a trusted source.

So, other comments in our group? Anything that we --

Okay, in terms of involving clinicians, hospitals, providers, getting their buy-in, again, the validity issue, face validity, scientific validity is very important here as well. You know there was discussion about the measures needing to be related to something within their control and

that they could improve. So, from the provider stakeholder perspective, this was what we discussed.

Again, earlier on this morning, we talked about the role of regulation or from a regulatory perspective. So, again, there could be value in and of itself of documenting variation in efficiency. You know you might say that this is regulation on the part of a payer, particular a public payer, but there may be regulatory decisions that could be made based on this information and it could be useful for incorporation for that purpose.

Payers, obviously, used for negotiation to avoid unintended consequences. And somebody may want to speak to that because I am blanking on sort of the point here, making sure that we are measuring what we think we are measuring and good communication message to explain these concepts.

Does somebody want to address the unintended consequence issue?

DR. RASK: The unintended consequence issue is just being sure what metric you are using for cost and that you don't do it in such a way that there is an unintended consequence of a provider looking good on that measure by denying access, by denying appropriate care.

DR. SCANLON: Okay. In terms of organizations, the ability, so how would organizations sort of internalize or use this, the ability to use and apply to their users, their audience, making sure the information is consistent and maintain integrity as it is disseminated. And this may lead to complex methods being lost or adulterated as they are passed on. And so that is certainly something that one would need to look out for.

In terms of what type of efficiency could be reported, we had, I think a fairly significant discussion on this is that discrete quality measures that actually represent efficiency, for example, things like

readmissions, and there was some discussion that in and of itself, preventable or avoidable readmissions are, by nature, potentially inefficient.

Discrete episodes that could be selected or improved or broader report card categories or stars for overall efficiency for clinicians or hospitals. And I think I will ask Iyah, do you want to say a few things on this? We had some discussion on this point.

MR. ROMM: So, I think the challenge that we wrestled with is that we felt that were in some ways treating all efficiency as the same and recognizing that there probably are many flavors of efficiency.

There is the vector of cost and quality which we spent a lot of time talking about this morning and/or the linkage of them side-by-side, even if they do not combine into some vector.

There is the efficiency that, as Dennis said, is sort of the quality measures

in and of themselves that may or may not be truly what we mean by efficiency but I think ease of meaning. And when we think about things like choosing wisely and the wasted space there, I think there potentially is import in bringing that back to bear as well and really highlighting that area in a consumer-oriented way around those things that both are inefficient and potentially harmful.

And then we have done a lot of work in recent years on things like segmenting certain episodes, deliveries, hip and knee replacements, other things that patients can make choice around.

I think one of the challenges, and back to your other comment around consumers want this information, I agree with you in every conversation that I have with consumers is in that direction. Part of the challenge that we face is in Massachusetts we had a cost and quality website for a number of years that had dollars signs and stars and nobody used

it. And it had every provider. It had physician groups. It had hospitals. It had all sorts of procedures. It had the same quality measures we are talking about here, side-by-side and it was never in use.

And so I think we have to face sort of a fundamental challenge. And we stopped doing it, Massachusetts as a commonwealth stopped doing it. So, there is a challenge that we were also wrestling with in this conversation around we all believe that this is the case but there is sort of an information and use gap that we have to figure out how to address in that public reporting space, specifically around consumers.

CO-CHAIR DUBOW: And my two-second answer: I think there is value to publishing, even if they are not using it.

MR. ROMM: Agree completely. I am not a proponent of not doing it. But I think that specifically for the consumer segment, I think that we all agree that there is value

simply to transparency and that the sunshine that comes along with that and the behaviors that potentially change, but specifically on a consumer use perspective and driving choice, one of the things that we spent a bit of time talking about is that we can talk a lot about quality but actually measuring cost and pivoting cost against quality is incredibly hard. We had some conversations about things like facility fees.

And within a given organization, the fact that different sites of care have various rate bases integrated, academic, affiliated, there is huge variation here. So, what cost means in this sort of a setting, I think is very complex.

DR. SCANLON: And so yes, I don't think message should be that we, as a group suggested that this shouldn't be valuable or there shouldn't be a use for consumers but there is a lot of unknowns and I think a lot more study and research needs to sort of go in

to sort of help how to understand where there is value for a consumer and how to present it.

So, I won't necessarily, in the interest of time read through this list, but you can see that the first two certainly do address that issue related to consumers. What do they want. What is efficiency from their perspective. Are the measure developers' notion of efficiency consistent with sort of how consumers or consumer audience might interpret that information?

Other important issues, such as should we normalize quality and look at price? What is quality worth? There is a whole host of statistical issues and things related to timing, decisions about do we standardize input prices or do we sort of allow for market variation in pricing and account for that in terms of our measures of efficiency. What is the right choice to make and does it depend on the particular situation for which we might be publicly reporting?

And I think that is it.

Let me pause, as well, and see if there is anything else from our group's perspective. Okay.

MR. BECKER: I just wanted to say that I am right where Joyce is. Put the information out. People will figure out what to do with it and how to use it. If Chester Carlson and Joe Wilson had listened to the first person that told them that the copier was going nowhere, where would we be? Or Steve Case, when they told him, this IM thing, that will never work. And they told Lonny Reisman the care considerations, he was a blithering idiot.

So, put the information out and let's see how people will use it and manage it and I think people will make use of it in good ways for consumers and others.

DR. SCANLON: And just related to that point, I think we talked a little bit about that and sort of said we have a recent

experience of that and one that is not necessarily too distant from this topic, which was the release of the information last week by CMS. But I think that some would argue that what you just described is perhaps what just happened. And of course if you have been paying attention to the news, there is a whole variety of opinions about that.

MR. BECKER: It is way too early. It is two weeks.

CO-CHAIR DUBOW: But it is premature to evaluate consumer use of information. I always talk about premature evaluation. You know the data are getting better. The efforts to produce the information and to present the information are increasing. But many people don't have access to it or know where it is.

This is new. And I think that we should not be hasty in evaluating it and in understanding who is using it. I just think that we should not jump to conclusions about

whether people will find this interesting or not.

MR. BECKER: In five years, just try to take an EMR away from a physician. Just wait five years.

DR. SILBER: Can I ask Iyah why do you think the consumers in Massachusetts didn't use the quality report? Could it be that they didn't believe the quality report or were there issues about the quality report? Was this the Shahinian model for surgery outcomes that was in that report?

MR. ROMM: So, could I take the liberty of the microphone and ask a question that we were asked sort of softly around our group? A show of hands. How many people in this room use quality data before you choose where to go to get your quality data.

Exactly. And so I think that --

CO-CHAIR DUBOW: Ask how many people have access to some information.

MR. ROMM: How many people feel

that you have access to some information?

DR. NEEDLEMAN: Well, there is some.

MR. ROMM: So, I think my point is I am absolutely with you. I think that the challenge that we faced in Massachusetts is that the information that we have in hand has proven not to be good, useful information or has proven to not be perceived as good useful information.

And I think that it also was not -- you know no marketing, no communications, no sort of outreach to consumers. But also know evaluation of how consumers are interested in using the data. And I think that that is a real lesson to be learned. And we are trying to go at it again. Massachusetts is working to build a new consumer website oriented in a different way. But interestingly, not oriented at data, oriented at how consumers make decisions in general.

So, I think that there is an interesting conversation to be built upon. I am a huge advocate of this sort of data transparency. I was a regulator for a while. But I think that our early lesson, and we had it for four years, it is still up, the data are just very old and are not updated anymore is that people the hits tracked in the hundreds on annualized basis and it was almost all researchers.

CO-CHAIR DUBOW: I don't mean -- we don't need to shoot the messenger. We appreciate your efforts.

DR. NEEDLEMAN: This issue of trying to understand how consumers make decisions and where information falls into it I think is central to this. If you are going to try to put this stuff out, it is not just how it is packaged.

I mean, look at Bill Clinton. Right? Bill Clinton goes to his doctor. He is told he needs by-pass surgery. New York

State has been publishing for a decade information about both hospitals and individual surgeons. He is clearly going to a good doctor because he has got access to a good doctor. New York Presbyterian Hospital, two different campuses, the Presbyterian campus, the New York Hospital campus. The New York Hospital campus scores right on the cardiothoracic surgery data and the Presbyterian campus scores poorly on it. And he winds up at the Presbyterian Hospital and ultimately winds up with a complication that is due to the poor surgery he got the first -- he got.

If he and the doctors serving him can't get it right, with all the data that is out there, what is going on? And until we figure out what is going on and how to change the decision-making, this data is going to go out there. It is mostly going to be an embarrassment to the providers and that is what is going to generate the change, not

consumer behavior.

DR. ASCH: Two things. One, just a response to this ongoing conversation. Everybody makes decisions that are socially constructed. I heard behavioral economics mentioned earlier. It is not like we just all look at numbers and come to some rational decision about it. I imagine Bill Clinton is no exception. When he was the President, I remember him making some pretty terrible personal decisions.

(Laughter.)

MR. ROMM: He didn't think so.

DR. ASCH: In the end, I bet he did think so.

And so I guess what I think is likely to happen for consumer choice on efficiency is that people will end up with narrative descriptions on the web like Yelp or Doctor Choice, there is a zillion of them out there. But the numbers that groups like this would probably be more comfortable with will

inform those narrative descriptions. Let's hope that is the way it will be. But that wasn't what I was going to say. That was just a reaction to what you just said.

What I was going to say is I am surprised that we have not mentioned measurement burden as one of the ways in which we should choose between measures of efficiency. We could actually adversely affect efficiency directly by requiring a lot of extra measurement. And if it was going to come up in one group, I thought it would come up in this one. Did it not?

MR. ROMM: I think sort of tangentially. But I think that most of the frame of the conversation was really around use of existing data to report on efficiency. There really wasn't conversation around generating new data. And I think that may be just sort of individualized perspectives.

From my state, I can't think of anything on the state basis that I would need

to generate to do any of the things that we talked about. It is a question of whether we are comfortable with the calculation.

You know things like cost accounting came up tangentially as sort of broader structural challenges around capturing efficiency as come up in the first group but we didn't really go down that pathway.

DR. ASCH: The only reason I bring it up again is that if we are really going to try and synchronize cost and quality measures for true measures of efficiency, as you brought up a few times and I agreed with, it may actually require changes in the way data are collected and that could impose a cost on the providers and health plans.

CO-CHAIR DUBOW: You know I meant to ask Andy and Chris earlier today about feasibility with respect to data collection and things of that sort among the approaches that they proposed, in terms of whether these are equally feasible, whether one requires a

whole lot more effort, whether the skills.

Somebody talked about -- Cathy it was you who talked about big insurers, small insurers, and whether there is the capacity to do this. And I think that is something that has to be taken into account.

DR. MAC LEAN: I agree, Steve. We do have to consider measurement burden.

I think, though, we need to get back to first principles. And I would hope that the charge of this group is to actually define what is needed to produce valid, meaningful, not just reliable, the difference between reliability and validity measures. And I think that part of the measurement burden that we currently have, quite honestly, is because we are trying to retrofit data that were never intended to be used for quality, for example, into this quality construct. And it is problematic.

So, I think that we ought to define what is needed and define what is

needed to get there. And hopefully, we can then move to that place.

MR. ROMM: I guess just a last thought that we spent a few minutes talking about that I think is worth us carrying forward. We spun back at the end of our conversation to the fact that I think our perception was the task of this group is really thinking then about how do we take all of these concepts about how we should be using efficiency measures and translate it to selecting measures themselves and found that much of our conversation really came back to either measures that exist in the form of not linking cost and quality but rather those measures of efficiency, case mix adjusted, average length of stay came up, readmissions, those sort of things that we can start to promote a little bit more, choosing wisely and the sort. But also then simply the side-by-side pairing.

And we talked about things like

ways that consumers themselves or other users of data could stratify up or down the various weighting, if you will, and whether there were ways that we could think about simplifying approaches to do such in a public space.

I think, though, that the idea of sort of then how do you select measures appropriately was something for public reporting specifically that we really couldn't make any headway on. And so I think that is sort of a broader challenge of we really don't know what is and is not appropriate in the space to put out publicly and that is something that needs a lot more explanation. My two cents, personally, is that we should put everything out.

CO-CHAIR DUBOW: I think we should move on because there is so many ways to respond to that, because I am looking at the clock now.

Group three is the network design tiering group. I don't know who the

spokesperson is.

MR. AMIN: I think actually I am going to take that for the group.

CO-CHAIR DUBOW: Really? We are not doing everything?

MR. AMIN: No, we are. I will take group three.

CO-CHAIR DUBOW: Oh.

MR. AMIN: I was the facilitator but I didn't assign somebody to report.

CO-CHAIR DUBOW: Okay.

MR. AMIN: We were going for quite some time.

CO-CHAIR DUBOW: And no slides.

MR. AMIN: No, there are no slides. We went all the way to the end. Is there a microphone still or no? Okay, thanks.

Okay, is this working? Okay, great.

So, again, I would welcome the input from the committee members -- from our group members on this.

So, ours was around network design and tiering. And the basic function of network design and tiering that we discussed was basically which providers are in the network. And there are three elements that were at play for this particular use case, the first being network adequacy, the second being the cost, those two being the major drivers, and then quality.

So, we discussed multiple different audiences for this use case, the employers one of the main and cost being a major factor for them. And then additionally, consumers, and then providers and then a regulatory.

So, when we were thinking about the question of looking at the link between cost and quality, we noted some general considerations that there really isn't full information about cost and quality, particularly around the indirect cost that consumers would likely face in consuming

healthcare. And particularly, we didn't want cost to be used as a proxy for quality. And then additionally, the concern around access, that we don't want to limit the number of providers, potentially, in particular markets.

And so when we looked at the question of complexity for the different audiences, in particular consumers and employers, for consumers we generally felt that there should be more of a straightforward approach and generally simple to understand, which generally lent itself to the side-by-side model that allows for ease of interpretability.

For employers, we discussed that there was probably a tolerance for something greater than that, although with the shift to exchanges, that would be a topic that I would welcome discussion from the committee, the small group, that we discuss that a little bit more.

So, from the health plan

perspective, what we basically discussed was that really there is more of -- the current state is sort of the cost hurdle and then a look at quality later, the cost being the negotiated price being the major driver and then quality standards are often part of the discussion but are sort of secondary to the first cost hurdle.

And then it is also complex, given that quality may vary across service lines for different accountable entities, in particular, for different hospitals or different service lines within hospitals. But generally the transparency may help to drive improvement.

And then finally, our major discussion that we sort of ended with was around scoring the difference between relative and absolute scoring, that in this particular use case, we may be better off creating thresholds slash hurdles because there may not be sufficient discrimination among providers. Although, there was a concern that this type

of approach might not incentivize improvement.

And I think that was the basic -- we also drew a distinction here between when we are looking at the various models, which considerations we would make for the methodology and then which considerations we would have for the display.

And for the methodology, really today it is a cost hurdle, as I described before for inclusion in the network and then quality comparisons. But the methodology should be fair and scientifically acceptable. And for the display, really that it should be understandable and actionable.

So, I welcome reflections from the committee about the discussions about any of those topics or any that I missed.

DR. RYAN: Taroon, I have a question about this absolute versus relative scoring.

It seems like the whole point of tiering is some relative distinction.

Providers are on different tiers. So, it seems like if everyone is on the same tier, everyone is good, then it seems like the tiers aren't doing anything, unless the tiers distinguish providers somehow and has a relative comparison to kind of not functioning in terms of how I see them.

So, I just wonder if there any comments from the group that could address that issue.

MR. AMIN: Yes, I would welcome comments on that topic from the workgroup.

DR. MAC LEAN: But there is two dimensions, right? There is both cost and quality. And that discussion that we had was about the absolute was really relative to the quality component. And we kind of expand -- so you could have equivalent quality and have different costs. Right? So, there could be differences there.

So, I think that should answer your question.

DR. SPEIR: I'll just mention on thing. I think there was also a consensus that perhaps that scoring system may also have to do with both the robust data with some providers and the paucity of data with others.

And so the necessity to score in a consistent way may be as much altered by what we are seeing to build across the board with what is currently present with some that have kept such data, like in my specialty, for a long period of time.

Secondly, is what is good enough. And as we are moving forth in this particular setting, particular thresholds, is it three tiers where you have got the very good, the good, and the not so good, and so is it good enough to be in the good category.

Or if you have got others that are over 90 percent in the scoring system, is there really any value in looking at whether somebody is at 91 percent and 93 percent? Is there the ability to differentiate in that

regard?

So, as you are building such scoring systems, particularly in provider networks, what types of categories would you build, five categories or three or one, or those that aren't any good at all?

And that led us then to the discussion about access, which is a real issue both around manpower, as well as when we start scoring, particularly paying for outcomes, our value. Are you going to be limiting a whole group of providers that aren't going to be acceptable really for providing care? Where does that take us?

So, those were some of the general ideas that we had some trouble getting clarity for. And it was very helpful to have both a cross-section of payers, as well as providers.

CO-CHAIR DUBOW: Steve, you had a -- yes?

DR. ASCH: Actually, you said mostly much of what I wanted to say. Great

job on doing the report out for a very talkative group. It's okay that you didn't tell one of us to be the reporter. I think you probably did a better job than any of us would have.

But the only thing I would add to what you just said, is to emphasize something you went over very quickly, which is there is a problem with all threshold measures, which is the incentive for improvement is limited to just getting over the threshold. Say you have a measure of quality and you have to get above a certain measure of quality and after that, only cost matters. Well, after that, only cost matters.

So, if you set it too low, then the incentive to do better on quality is limited, which is why I was very glad to hear you talk about the number of levels that might be needed for a system of measuring value or efficiency.

CO-CHAIR DUBOW: Steve, why can't

you keep raising the threshold?

DR. ASCH: You could keep raising the threshold but you are only incentivizing innovation at the bottom then. You are not incentivizing people to continuously improve. And the quality improvement literature, at least to me, and there are a lot of people around this table that know it as well as I do would lead you to believe that that is not the best way forward.

CO-CHAIR DUBOW: What about changing the -- swapping out the measures once the desired -- when they are topped out?

DR. ASCH: I'm all for that but I think the same criticism applies to all threshold measures.

CO-CHAIR DUBOW: Okay. I think Jeff is on the group. So, just -- oh, no, you are on the other one. Never mind. So, Greg first. You are in the same group.

DR. WOZNIAK: Yes, maybe this is more of just a clarifying question. People

probably can't see it but on the bottom sheet there is creating a link between cost and quality. And then there is full information and then there is indirect cost. What were those indirect costs? And are those indirect costs to the plans, the consumers? Another cost concept that just came up.

MR. AMIN: If you can speak to it, I would welcome the small group. But I think the way that it was referenced was in particularly related to interactive -- you know the example that we used was if you created a network where it might be cheaper to send your patients to Cleveland but there is additional costs related to time away from work to be able to do that, your family has to fly across the country, those costs that are potentially borne by the patient that might look efficient from a cost perspective. But their costs are borne by the patient and may not be really captured in the system. But again, I would open it to the work group.

DR. SILBER: I wanted to explore the idea of the hurdle again, for what you were talking about.

So, if you reach a certain hurdle, a certain level, then you can use the efficiency metric. I think that is what you are saying. But below the hurdle or that you can't, because you wouldn't say if you want quality you would be too low.

But so what is really happening is unless you use your metric along with a side-by-side, you are saying after we get to a certain level of quality, we can ignore the data on cost and quality and just use the efficiency metric. Am I right in suggesting that is what you are suggesting? Because to me, that would be a mistake. The concept of a hurdle is a mistake and it should be you should see the cost, you should see the quality. And the hurdle is confusing things because you are depending on the efficiency metric after the hurdle when it is still

relevant to know what the X and Y is, what is the cost and what is the quality.

Maybe I am wrong on how the hurdle is being used. So, I just wanted to understand that.

DR. LOWE: It wasn't the -- it was when you reached a certain threshold of quality, then they compete on price. Because once you -- this is an acceptable threshold. All things being equal, if physician A, physician B, physician C all produced the same basic outcomes, then what differentiates those three positions is pricing, not quality.

DR. SILBER: So it is a huge assumption to say that they are the same outcomes. Once you are beyond a very bad level -- above the minimum and then you are all the same is a bit odd. Even dividing it up into five groups, Cochrane would say five is generally a good number but that is because he said you could do statistical adjustments after the five groups. That is the famous

Cochrane paper, right? We are not doing that. So, even five isn't right.

So again, if we are going to get into this point where we are going to say well, we can use efficiency after we get to a certain point of quality, then I think it is very dangerous and I would say that we really should think hard about advising that that could be an acceptable metric.

DR. LOWE: I wouldn't call it efficiency. It is cost.

DR. SILBER: Well, okay, it is even worse.

DR. LOWE: The assumption would be as cost went up, efficiency would go down.

DR. SILBER: So you are really, really in dangerous territory when you are then assuming that say two-thirds of the quality is all the same and we are just going to go for the lowest cost. That is really dangerous. That is exactly what we wouldn't want. I wouldn't want for the NQF to endorse.

CO-CHAIR DUBOW: There are a lot of signs up and I don't know if -- Cathy, do you have anything? And Steve are you and Andy and --

DR. MAC LEAN: Just commenting further on I will answer that and then make the comment.

I think that the whole discussion around efficiency or value, I think the first thing we called out in our group was that it is difficult to kind of in this sort of a network sort of design to come up with a single metric. And in all cases we, I think, advised a side-by-side model.

So that you might have some threshold to get into it but that in all cases you would be reporting out kind of side-by-side what the cost is and what the quality is. Because the whole thing is confounded by so many factors with regard to getting into the network and negotiating the price with the providers. It is all about the negotiation

and it can get quite complicated, those negotiations. It is not just no, I want more money. And kind of highlighting my point, it can even get to the point where well, okay, I will give you this for this procedure. We will give you this price for this procedure but you have got to make it up on this one. Or, okay, I will negotiate that with you but only if you let me -- say, we are talking about a commercial contract -- I will do that for your PPO or HMO business but you need to put me into your contract for the exchange business.

So, the price piece is very complicated with regard to what you are actually paying. So, that is in response to your comment.

The other thing I wanted to raise, was we were getting into a very interesting discussion when we had to come back here and why we didn't have slides. And Steve was really kind of leading this discussion. I had

this raised this on one of our prior calls. Is the concept of how do you value -- what is the value of a procedure in different populations or what is the value of different procedures across different populations? So, the example that was given in our group was doing a valve replacement in a 95-year-old versus a valve replacement in a younger patient. What is the value to that and how can we concur that? And Steve brought up the concept of quality adjusted life years. And I think that that is important, not only in comparing the value of doing the procedure within -- as single procedure across different populations but also comparing the value of different procedures. So, a total knee replacement in a 65-year-old, versus the third stage of chemotherapy in someone who has got metastatic pancreatic cancer.

So, I think that -- I don't know if we should explore that more but I think that this quality adjusted life year concept

or some other sort of concept of outcomes we can measure across patients and populations is worthy.

DR. ASCH: Can I respond to that directly or is there somebody else in line?

CO-CHAIR DUBOW: Yes, do it but -- yes, please.

DR. ASCH: Last comment. First of all, thanks for tagging me with the death panel thing. No problem.

(Laughter.)

DR. ASCH: In all seriousness, what this means to me is that we have to have ways of comparing across different procedures and across different clinical situations. And I don't know, I was mentioning to somebody at the break, the article that I really enjoyed by Steve Schroeder, putting the value back in the relative value unit. Did anybody get a chance to see it? It was in the New England Journal a while back.

So, the idea is that if you pay

for something that is by volume, like we do now, we get what we have now, which is a dysfunctional system. But on other hand, there is a modest tweak, at least, that we could do. And I think health plans could lead the way here, along with the government. And that is pay for appropriate volume, pay for volume that is likely to lead to a benefit, even if we don't do formal cost effectiveness analyses and actually value people's future years in dollars. I think we could get part of the way in so doing.

CO-CHAIR DUBOW: And then we are going to go to the last group.

MR. ROMM: I guess just a quick comment back to the points that Cathy was making around this issue.

I think that one of the really big challenges here is that I would posit that this actually is not a question about efficiency measures at all but about what the role of the NQF in this process is in thinking

about sort of the principles of how these measures come together in limited and tiered products. Because at the end of the day, the challenge here is that not only is there a price negotiation the way that you are talking about but people are negotiating what quality measures are in.

So, I know of limited networks where the only quality measures that are captured are five HEDIS measures where everybody looks the same and a bunch of Hospital Compare measures where everybody looks the same. And so that is not to say across all quality measures everybody looks the same, but rather to say that where it is not only a negotiation of base price but a negotiation of what measures and people can therefore game that element, too.

There is actually a really interesting space for a policy conversation. I don't know if this is the right setting for that, though and I feel like we could veer

down that path very far.

CO-CHAIR DUBOW: Well but the issue of -- I mean I don't know why the policy question is not relevant to the NQF process of providing meaningful and useful information to people who may be directed to one tier, one network or another, depending on how it is constructed. And so, giving guidance as to what is methodologically fair, for example, to the range of users, providers, consumers, may be quite relevant to the NQF process.

MR. ROMM: So, I can't speak to that. I would love to see it. I would love to see somebody who is really taking that ball and running with it. I think it is a huge opportunity and a huge gap in the way that we are thinking about these products popping up all over the place. I think that the biggest threat to them is this negotiation question. And market leverage comes into bear everywhere but it is really starting to intrude into a product that was designed expressly for the

purpose of overcoming that market force. So, I would love to see that. I just am not sure what the right setting is.

CO-CHAIR DUBOW: I am simply suggesting that, as a minimum, we need some outward bounds. It may be one of our audience's regulators. Maybe this needs to be regulated because it has that kind of impact on individuals. So, just a thought.

We have one more group. We are going to just bleed a little bit past five o'clock, since you are all coming back tomorrow anyway. So, we figured we would take advantage of that.

And the last group is the P for P, the pay for performance.

MS. WILBON: So, Gary is going to be our spokesperson. Thanks, Gary.

DR. YOUNG: Okay, so I will go quickly. I will speak quickly and I will also outline what I thought was a fairly wide-reaching, far-reaching discussion. And I will

also invite my group members to sort of amplify what I am saying here, as well as correct any misstatements that I may make.

So, this was for pay for performance. We saw our audiences, including purchasers' plans and also consumers, because they ultimately get affected by the design and implementation of pay for performance programs.

We saw as an important principle that pay for performance programs should not just entail a single efficiency score that would be used for payment purposes and for communicating information to providers but that it should also include side-by-side information that providers should be given side-by-side information for cost and quality and let them have that information as part of a pay for performance program.

Also that payers certainly should reward providers who are at the highest levels of quality and the lowest levels and at the

lowest levels of cost. So it is sort of in that top quadrant there. And at the same time, the providers should be rewarded who are at the high end of quality but where there is still opportunity to reduce costs.

More debate over whether providers should reward, whether payers should reward providers for improvement for those, for example, at the frontiers of highest efficiency but low quality, whether there should be rewards that are made available to those who sort of improve when they are already at the frontier for high efficiency but also low quality.

In terms of scientific rigor, there was agreement that whatever methods, models, approaches that are used, we want to see the rankings, the relative position of providers, that that be robust to different models, approaches, and that there is robustness there and that a provider's ranking or standing isn't sensitive to one particular

model or approach, that it is robust across models and approaches.

In addition, there was discussion and consideration about including outcomes in pay for performance approaches in that if you are including outcomes, there should be some concern about possibly using indirect adjustment approaches. When you have differences in population characteristics among providers and that the more different those population characteristics are, the less they overlap, indirect standardization becomes increasingly problematic. And so then you need to be thinking about direct standardization, again, if those population characteristics really don't overlap in a substantial way.

Additionally, another consideration around scientific rigor is the use of standardized prices and a recognition that standardized prices become very problematic because of cost shifting and that

we have got different providers with different percentages of their patient population that are uninsured, Medicaid, Medicare, and private sector, standardized prices can become very problematic because you have got providers sort of shifting those costs onto the private sector side.

And so from a resource perspective, they are looking more expensive than they may actually be relative to those that have patient populations that are largely private pay.

So, I will turn it over to my group members to add anything.

CO-CHAIR DUBOW: Anybody in that group have something to add? Alan, you have a question or comment?

DR. SPEIR: I just have one comment. We have a nine-year history of a pay for performance model in Virginia with Anthem for the cardiac surgery population, where it was both process as well as clinical outcomes

that was normalized with observed or expected complications and mortality. And there was not an absolute dollar amount, rather a percentage over what was normally reimbursed to the physicians.

CO-CHAIR DUBOW: Any other comments or observations? Everybody is very tired now. We are winding down.

You know, Andy and Chris, I don't know whether you want to give us your reactions now or wait until the morning, until you have a chance to think about this.

DR. RYAN: I have a 45-minute prepared statement.

(Laughter.)

DR. RYAN: No, I have nothing more to add right now, Joyce.

DR. PANTILAT: You can turn that into a PowerPoint.

DR. TOMPKINS: I also have to defer my 45-minute comments. I have a call that starts at 5:15 and I don't get coverage

in this building, so I have to go down the elevator.

MS. WILBON: On that note, I know everyone is tired and this will be perfect timing to go two blocks away called Mio and help you unwind.

So, we will wrap up now. I think we have decided that we are going to do the actual Day 1 recap tomorrow. It will give us all some time to absorb and I am sure we will all come back with some more ideas tomorrow on kind of where we landed today on our discussion.

So, for tomorrow, we start at the same time. Breakfast starts at 9:30. We will begin the meeting at 9:00 a.m. We will shift some things around a little bit, since we actually finished the report outs of the break out groups early. So, we will probably shift everything up and I don't think we will have a problem filling the time tomorrow.

So, tomorrow morning when we get

started, we will kind of give an overview of the agenda and where we landed with things.

Unless there is any other kind of parting -- oh, thank you! Public comment. Thank you.

Is there anyone in the room or on the phone, Operator, who would like to provide a comment to the panel?

OPERATOR: At this time, if you would like to make a comment, please press * then the number 1.

There are no comments at this time.

MS. WILBON: Okay, sound like we're -- everyone is just worn down. So we are going to let you go.

We will recap the agenda for tomorrow and thanks again everyone for a great day of really great discussions.

(Whereupon, at 5:07, the meeting was adjourned to reconvene at 9:00 a.m. on Friday, May 2, 2014.)

C

D

E

F

J

O

P

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: 05-01-14

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