

NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
October 11, 2012
12:00 p.m. ET

Operator: Welcome to the conference. Please note today's call is being recorded.
Please standby.

Elisa Munthali: Thank you, (Raquel). Hello everyone. My name is Elisa Munthali, and I'm a senior project manager with the National Quality Forum. I'd like to welcome and thank you so much for joining us today for the Composite Measure Evaluation Guidance Project. And today, we are trying to give an orientation to the technical expert panel. Also working on the project are my colleagues, Karen Johnson and Karen Pace. Both of them are senior project managers – sorry, senior directors on the project.

Before we continue and go through today's agenda, we did want to do a quick roll call and because this is our first meeting, we wanted to make sure that each technical expert gets an opportunity to introduce themselves and just give a little bit of background what affiliation you have and your experience working with composites or relation to this project.

So, we'll go down the list, it's an alphabetical order. And we'll start with Dr. Birkmeyer.

John Birkmeyer: Hello, everybody. John Birkmeyer, I'm a surgeon and my – I'm director of the Center for Healthcare Outcomes and Policy at the University of Michigan. My experience with composite is two folds. First, I'm part of the so-called "Blue Ribbon Expert Panel for the Leapfrog Group" which four months ago went live with its first iteration of hospital-wide composite measure of safety; and number two, with my collaborators, Doug Staiger and Justin Dimick, I

collaborated on empirical-based composite measures of mortality for inpatient surgery.

Elisa Munthali: Thank you, Dr. Birkmeyer.

Is Dr. Chase on the line?

James Chase: Yes, this is Jim Chase, and I'm president with Minnesota Community Measurement. We're measure stewards for two NQF-endorsed composite measures, are optimal diabetes care and optimal vascular care. Both of those measures are used across the State of Minnesota with about 400 practice sites of care across the state and are publicly reported.

We've also done some couple of other composites. One was a cancer screening composite, but we don't report that any longer, but glad to be part of the committee.

Elisa Munthali: Thank you very much.

Dr. Sherrie Kaplan?

Sherrie Kaplan: Hi, I'm Sherrie Kaplan. I'm the psychometrician by training. I'm assistant vice chancellor for Healthcare Measurement and Evaluation at U.C. Irvine. My experience with composite measures as I've – we've created some composite measures for physical-level for diabetes performance for the DPRP Program of NCQA working currently with this and a number of other composite measures sort of things on reliability of single versus composite measures with Tim Hofer and others.

I'm currently working with the State of California in their Safety Net Institute on nested sampling for composite measures for physical-level performance.

Elisa Munthali: Thank you, Dr. Kaplan. Lyn Paget?

Lyn Paget: Hi, my name is Lyn Paget, and I assume I am here more representing the interest of consumers and patients. I have recently started in the independent collaboration of patient policy advocates called "Health Policy Partners."

Prior to that, I was for seven years policy director at the Informed Medical Decisions Foundation in Boston.

Most of my measurement experience would come from not only being involved in policy development and work at the National Quality Forum, but I also was instrumental in the establishment of Medical Outcomes Trust which was a follow-up to the Medical Outcomes Studies to take patient-based, patient-reported outcome measures of functional health status and other health-related quality of life and put them into public domain for use around the world and I probably know just a little bit about composite measures have read everything that was sent but I'm happy to learn, and I hope I can be helpful to represent the interest of patients.

Elisa Munthali: Thank you very much, Lyn. Dr. Steven Wright?

Steven Wright?

Steven Wright: Hi, I'm a Health Services researcher by training. I'm with the Department of Veterans Affairs specifically within the Office of Analytics and Business Intelligence Office of Performance Measurement. My responsibilities include directing extensive measure management system for that performance measures in the V.A. and within that inventory is a vast array of composite measures. So I have a lot of experience working with different types of composites and different kinds of settings but, you know, particularly within a complex healthcare setting at the Department of Veterans Affairs.

Elisa Munthali: Thank you very much, Dr. Wright.

And Dr. Alan Zaslavsky?

Alan Zaslavsky: This is Alan Zaslavsky; I'm in the Department of Healthcare Policy at Harvard Medical School. A lot of my work in quality measurement has been with the Cancer surveys. I've involved in quite a number of different development and implementation activities with (CAP) which is reported in composites. I've also done some work on composites involving (CAPs) and (inaudible) I'm a statistician by training and with a lot of focus on quality measurement over the recent years.

Elisa Munthali: Thank you very much. And I just wanted to mention that Dr. Dale Bratzler and Dr. Elizabeth DeLong unfortunately couldn't join us today, but there will be with us here in November at our in-person meeting.

And I also wanted to remind the committee that the proposed roster with your name, recommendation to be on the technical expert panel was posted to our website for 14-day member and public comment and so as of Tuesday night, you are officially on the technical expert panel, but as a result of that process we received additional recommendation, three additional names to be placed on the technical expert panel.

Those are being vetted internally. They're going through review by our Senior Management and our Consensus Standard Approval Committee, and I just wanted to recognize any of the three members that are here so they can also introduce themselves and also – yes, so that they can also introduce themselves. I am not sure if David Shahian is on. I think he is. I see him online.

David Shahian: Yes, this is Dave Shahian. I served on the Initial Composite Steering Committee at the NQF. I'm in MGH in Harvard. I chair the STS National Database and its Quality Measurement Task Force. We've developed now two composite measures, one is already been in NQF endorsed for CABG. It's a multidimensional composite measure that we publicly report and then our second composite on isolated aortic valve replacement is about to be published in the peer review literature and will be brought to NQF next cycle. So, hope that I'll be able to participate. Thank you.

Elisa Munthali: Excellent. Thank you very much. And I'm not sure if Nancy Dunton is on the line.

Nancy Dunton: Yes, I am.

Elisa Munthali: Hi, Nancy.

Nancy Dunton: Good morning. My name is Nancy Dunton. I'm a research professor at the University of Kansas Medical Center School Of Nursing, and I'm the director

of the American Nurses Associations National Database of Nursing Quality Indicators.

I have no – we've developed no measures that have been endorsed by NQF, but I have two activities that are relevant. One is that we annually create a composite measure of nursing quality for awards to hospitals participating in NDNQI and the second is we've done some methodological work with Dr. Byron Gajewski in controlling for regression to the mean in composite measures.

Elisa Munthali: Thank you very much, and I'm not sure if Dr. Goldstein is on the line. She had a conflict and was going to try and make it today. OK and we'll make sure we get the updated roster to you as soon as the (addition) names are approved.

And so at this point, I'm going to turn it over to Karen Pace who would start the presentation.

Patrick Romano: I'm sorry, could I introduce myself?

Karen Pace: Yes. I just what – I was just going to check whether Dr. Romano is on. Is that you, Patrick?

Patrick Romano: Yes, it is.

Karen Pace: OK. You want a ...

Patrick Romano: Hi, everyone. Yes, this is Patrick Romano. I'm a general internist and a general pediatrician by training. I'm based at U.C. Davis Health Systems School of Medicine in Sacramento, California. I know I've worked with many of you before in various NQF activities and quality measurement activities. I've been involved in quality measurement for about 20 years mostly using hospital administrative data sets although we've done a little bit of work with other types of data as well, and I have been a part of the team that works with the agency for Healthcare Research and Quality on developing testing and validating the AHRQ quality indicators. So, in that

capacity, I've had some input into the development and now the revision of the AHRQ Q.I. Composites three of which are currently endorsed by NQF.

Karen Pace: OK. Thank you. And I'd also like to recognize and thank Dr. Romano has agreed to be one of our co-chairs for this expert panel, and we'll be adding another co-chair prior to the meeting – the in-person meeting, but I wanted to thank you for that, and we'll go on with the agenda and I'll turn it over to Karen Johnson and then for quick overview of the project in NQF and then we'll get into the – some of the more meaty things that we want to cover today.

Karen Johnson: Thank you, Karen. So, as Karen just mentioned, we will give you just a very brief overview of NQF. I know most of you are already familiar with us but we thought we should take the opportunity to do that, then we'll talk very briefly about the project including the timelines and that sort of thing and then we will talk about the things that we really need to hear from you today, key questions and issues, key references and then hopefully have some discussion about approaches and resources that you think that would help us to facilitate our discussion in November and then finally our wrap-up will be just housekeeping details about the rest of the project.

So, with that, what is NQF? It is a private, nonprofit voluntary consensus standard organization. It is a public-private partnership, and it's also a member organization. We have about 400 member organizations and they represent a variety of stakeholders including consumers and purchasers, health providers, health plans, community public health agencies and also the researchers and methodologist and suppliers and industry. So, we have a very varied stakeholder group.

Our mission – basically we have a three-pronged mission. First is building consensus on national priorities and goal. The second is endorsing national consensus standards and then finally promoting the attainment of national goal through education and outreach program. So, kind of a very detailed group in the middle there and then our inward and outward looks on the other end.

Our roles basically, we have two main roles – one is we are a standard setting organization, and this is then pretty much our foundational work since our inception. We endorse voluntary consensus standards for performance measures, SREs, preferred practices and framework, but more recently we have taken on the role of being a neutral convener of multi-stakeholder group and that really is done through (to) of our main works in the National Priorities Partnership and the measure applications partnership.

The NPP is a partnership to offer support and (consulting) service to HHS to help them set National Priorities and Goals for the National Quality Strategy and the National Quality Strategy is really just the our national blueprint for quality of health and health care in the U.S. So, right now we are – it turns out we are now 51 member organizations as far as the NPP and that work came to us through ACA, and also through the ACA, is our work with the measure application partnership. This is also another multi-stakeholder partnership created for the purpose of providing input to HHS on selection of performance measures for public reporting and also for payment programs. So, there's much more information about those on our website, and I believe we've provided those link to you if you want to read more about those.

And then finally, I think I wanted to just talk real quickly about why NQF endorsement and why it's important? Hopefully, everybody agrees that we need standardized measures so that we can assess quality and then hopefully improve it but also NQF endorsement really means that we have had experts – multi-stakeholder experts give a rigorous review of our potential measures of scientific and evident and really bring in lot of different prospectus from the different stakeholders so that the measures that we endorsed really are they have a – they reach a fairly high bar in order to look at measures for accountability and for quality improvement purposes.

So and that really (fits) in to what we're doing here. We know that our measure evaluation criteria has evolved over time as has our guidance that we've provided to those who are doing the evaluation, and hence, we need to think about the update to our composite guidance.

This next slide really is just a reminder to us that our national quality strategy aims for better affordable care and also better health for people and communities. So, that's always kind of in the back of our mind as we do our work.

So, real quickly I'll pause and see if anyone has any questions about NQF or our mission.

OK. If not, we'll go on, and I'll give you just very brief background for the project, as many of you know because you were involved in 2008 through 2009 NQF convened a technical expert panel to identify a framework for evaluating composite measures and as part of that work, a definition was developed, principles were articulated to (under why) –what needs to be done in terms of evaluating composite measures and actually specific criteria were developed and those fall within NQF standard evaluation criteria and as you probably know we have four main criteria importance to measure and report, scientific acceptability, usability and feasibility.

So, there are specific criteria under each of those that are specifically for composite measures. So, right now, we have identified approximately 35 composite measures that have been submitted to our – to NQF over the last what – three years since 2009 I guess, and of these, 21 has been endorsed. I have a little (inaudible) of line there on the 35 because some measures were actually submitted to us before our submission tool was online and that sort of thing.

So, doing the count should be quite simple and actually hasn't been quite as simple as we would have liked. But the purpose of the project is I'm sure you already know is to basically review and update our current guidance on evaluating composite measures and what we need to do as part of this project is basically realigned the unique composite criteria that we've already identified with an updated – with our updated guidance on both testing and evidence, and we did update the guidance for those two criteria I think in 2010, 2011. So, those have been out for a little while now.

And I think another driver of this project is that we have found – and Karen will go into this a lot more, but we've had inconsistent submissions and also inconsistent evaluation to some extent for our composite measures.

And then finally, our project activities and timeline – today is our first call where we're taking advantage of your expertise to help us plan a little bit for our in-person meeting and then after our in-person meeting, we'll have a post-meeting call on the 15th of November. So, our in-person meeting is on a Friday, November the 2nd, conference all after that on the 15. We will draft the report and since that's out for a 30-day comment period beginning right after Thanksgiving on November 28 and what we will do is actually bring our draft comment report to the CSAC in December and then right after comments have been gathered and received, we will bring you guys together again to discuss those comments on January the 3rd and then we will take final recommendations to the CSAC in January. And if all goes well, we're planning on board ratification by the 21st of January.

So, with that, I will ask if there's anyone have any questions before we get into the really (sun) part of the call.

OK. I'm handing it over now to Karen Pace.

Karen Pace:

So, again, thank you all for joining us today and helping us tackle this. As Karen mentioned that some of you were involved in the initial project. You know, we've had some experience now with composite measures in terms of their submissions and review in steering committees and we've certainly identified some issues and I'm sure you have as well. If you've been following any of these and so we really – as Karen mentioned, we've had some updated guidance in general about our criteria, about evidence in measured testing and, you know, we certainly need to be aligned with that but also to revisit these unique criteria for composite measures in terms of their testing and analysis and how these – how those apply.

So, I'm going to – and to just, again, reiterate what Karen has said, the purpose of our call today is not so much to resolve these issues but we really wanted to take advantage of these call to make sure – to help us with the

planning the in-person meeting to make sure we've got all the issues on the table to get your suggestions if there are additional resources that will help staff and you prepare to tackle these issues at the in-person meeting or any suggestions for the approach. We should structure that in-person meeting so that we, you know, one we have the one meeting – in-person meeting time, and we wanted to as effective and efficient as possible.

So, I'm going to go through the issues that we identified in your briefing memo. I'll try to touch on these pretty quickly and probably what we can do is I'll just stop and see if there's any needs for clarifications about this issue but the key things we'd like to talk about today certainly if there are additional things about these issues or any issues that we haven't identified that we want you to be sure to bring those to our attention.

I think we have a small enough people on this call. So – and your lines are open. So, you know, feel free to, you know, break-in or stop, but I'll to stop after each issue just to make sure that if we need to do any clarification (of why) it'd be easier to do it as we're going along rather than holding that.

So, one of the first things, and I think this has a cascading effect on all the following issues is that there's really been some confusion about what is a composite measure. At the – in the prior project, we've developed the definition of composite measure which is a combination of two or more individual measures and a single measure that results in single score and that was purposely developed to be pretty broad because there is a recognition, there were various types of ways of combining individual measure.

We also acknowledge that there were all or none composites, but I think as we've had some experience with these submissions it's become clear that we need some better clarification of what should be considered a composite measure versus what is should be looked at in different ways. So, I'm going to mention some of these things we've identified that has created some confusion. So, the first one is about – and I should say that the composites that we've endorsed have either been are all or none composites or composites which are actually some combination of separate individual performance measures.

But we've had some questions about our all-or-none measures for example because those are all done where the data are aggregated at the patient level. We've had – some of the submissions coming in clearly stating that their composites, others are simply submitted as a measure that has multiple conditions to be considered meeting the numerator conditions. And so, we need to make some decisions about how we classify those versus – and what the requirements would be if they are more similar to just a regular performance measure or whether they really a composite and require some unique analysis and testing.

And kind of along those lines, is that we have all-or-none measures submitted as composites but if you think of the reverse of that, any-or-all measures. So, an example would be a measure that is set up that if the patient has any of multiple lists of complications. Those typically have not even been submitted to us as composites. They're not viewed as composites by those who have been developing them but, you know, in that light should those be considered the opposite of all-or-none composite measures.

We've had measures referred to as composites that actually didn't meet our definition and I think, you know, one – and example of this and we can have some discussion about whether our ultimate decision of not classifying it as a composite is meeting our definition was correct, but the Leapfrog measure that John Birkmeyer mentioned in (terms) of them referring to it as a composite where they use volume as a way of their – part of their hierarchical modeling and their estimating and how they would target their shrinkage back to an average based on the volume or the size of the provider.

So, it didn't really fit kind of our definition of two individual measures but, you know, that's again, something we want to revisit. Something that was called out in the prior report is that paired or group measures that are just submitted to be reported together are not considered composites and, again, the other thing that we've had confusion about are measures that are based on surveys where the underlying instrument might be considered a composite but the performance measures themselves are individual measures. An example of that are (CAPs). The (CAPs) are based on a survey that has different scales

or domains but those are then translated into individual performance measures. I know that there has been some work done to actually been combined those individual domain scores into a composite but we've had some confusion on NQFs and of what those should be classified as well.

So, I think that's the first thing as we want to tackle what is the composite measure, you know, do we need to revisit that definition, we just need to get some clarity, and I'm going to stop there before I go on to the next issue just to see if anyone has any questions or comments about that or another way of looking that.

Male: One question I have what's the implications of that (part) of patient there obviously that's (inaudible) with patient. So, for example, if just wanted the all-or-none measure wasn't the composite or was the composite with that – that's the way it was handled in the (accomplishment).

Karen Pace: Right. So, that's a good question. What are the implications? And I think the major thing is that when with have – we had identified some additional criteria that needed to be applied to composite measures, some additional analysis and so – and also how we ask the developers to submit the information based on those additional criteria that are supposed to be met. So, it has the implications in terms of what testing is expected of those types of measures when they're submitted and then how the steering committee would evaluate as meeting our criteria.

So, with our current situation of some being classified a composites, some not, we've had some unevenness in applying those criteria that were identified in the original work.

Male: So, conceivably, we could come out with a recommendation that there should be more categories – one for (composite) – one for composites that are combinations of individual measures (consolidated) separately. Another one for – it's like all-or-none and another one – but then just one specific item.

Karen Pace: Right. That's a possibility and I think the other thing that really wants to do is we, you know, we have our basic NQF measure evaluation criteria, and if those were perfectly well, for example, with an all-or-none composite then we

don't really need to have separate criteria for them. We just say that, you know, they're submitted on our (regular) submission form and all those criteria apply.

So, you know, we have some different options, but you're right the first thing is to really sort out whether they are different and should have different requirements.

Sherrie Kaplan: This is Sherrie. I (cautious) – we're varied at – from a very different kind of perspective and psychometrically when you're trying to represent a conflicts construct like quality of care, even if you're doing it at the micro level just like (math) or other kinds of multiple complex constructs. Individual items are notoriously unreliable for doing that. If I ask you one question about your whole (math ability), you'd be unhappy.

So, the – for me the first step is to figure out how many items you need to represent whatever it is you're trying to represent and then the scoring issue comes after that. So, scoring is like a subset of how do you sample from all the domain of observable. So, things you could have used to represent that construct, to represent it fairly and then you tackle about the trade-offs between items per construct versus patients per item et cetera, et cetera and then you get to precision targets and then you get the scoring issue.

So, the all-or-none thing for example, if you have a nine item measure of diabetes performance and you say I have to (get) along of those nine items or I don't get paid. That's kind of separate issue for somebody like me than it is for obviously when you're looking at all cost readmission, you know, yes or no and that's a single item.

Karen Pace: Good. Thank you.

Patrick Romano: Yes, this is Patrick. I'm sorry, I haven't been able to get access to the SharePoint site yet, but I wonder if you could comment. So, I like this approach to dealing with the all-or-none composite problem because like Sherrie, I think I see it as a fundamentally different type of measure. To me, it's not really a composite measure at all. It's a checklist – it's a single measure that is labeled as a composite.

So, has NQF been consistent in the past in how it treats these all-or-none type composites or – and has there been some inconsistencies?

Karen Pace: There has been some inconsistency and that's why one of the, you know, that's why I'm laying out this issue about we need to really have some clarification and some precision about what is or is not going to be considered a composite.

So, for example, and you know, sometimes we have individual measures that maybe have just two conditions and, you know, for example assess immunization status and give vaccine. And those have not been submitted or viewed as composites but it seems that once you get past to conditions to meet the numerator people start thinking of composite and so we definitely have some inconsistencies and that's, you know, one of the things we really want to try to address and really get this sorted out so that we are moved forward in a more consistent way and really have the right criteria to be applied to the various types of measures.

And I'm sorry about the access to the SharePoint. We will, you know, whenever that happens if you let us know right away, we can get that resolved or send you things outside of SharePoint and so, if anyone else on the panel has had problems, please let us know right away and we'll make sure that we get that resolved.

Patrick Romano: The other thing I was going to say that I wonder if you could elaborate a little bit more on this idea and John may have other thoughts here about the Leapfrog composites. So, in this case you do have a measure. It's an individual performance measure that's a volume measure, and you have a separate measure which is a mortality measure, and they are being composite, right? Volume measure is being used to find the shrinking point in the hierarchical model but I guess I'm missing why that's not a composite.

Karen Pace: Well, go ahead, John.

John Birkmeyer: Yes. This is John. I wanted to weigh on that as well but also to be clear. You know, when I introduced myself earlier and talked about my role with

composites for the Leapfrog Group, I wasn't even thinking about the example that from a couple of years ago that Patrick just raised now, I was talking about the hospital safety composite ...

Karen Pace: Right.

John Birkmeyer: ... measure that both Patrick and I were part of. With regard to that – to that former measure, however, you know, I'll tell you first that I have no skin in the game with regard to whether it's classified as a composite or not. But, you know, but I sort of agree with Patrick that based on my pretty deep understanding about how that process works and what we're actually doing, and based on the simple but pretty unambiguous definition of what (own) composite is by NQF, it's hard for me to appreciate why you would not classify it as some (own) composite measure. It's taking too fundamentally different measures empirically combining them in a way that results in a single score.

Karen Pace: Right. And we can certainly get in to more depths on this, you know, in the in-person meeting and that's part of the reason we need to clarify it. I think the reason that we did not at the time is because you weren't simply combining – taking a volume measure and what that volume measure is and combining it with more mortality. You are using volume as you already pointed out to find the shrinkage point, and so the question, you know, and so it gets more in the whole hierarchical methodology in what you choose as your shrinkage and I – that's the reason whether that was right or wrong.

I think, you know, we're open to discussion about that but that was the reasoning at the time that it wasn't merely taking the volume measure and combining it in some way with the mortality measure. It was, you know, involved in the whole methodology. So, you know, we can certainly, you know, talk about that but again that has implications of what we would have required be submitted about the composite.

So, there's pros and cons to that, and we just need to figure out what's the right way to proceed.

John Birkmeyer: Yes, and – I – this is John. I totally agree with you that this is a topic for elaboration at the face-to-face meeting but that – but that particular application of the methodology that got put forward by Barbara Rudolph, Leapfrog group what sort of they (hyper) simplified sort of two variable composite approach to getting to that theme. In result, that methodology, you know, has been applied more commonly at least by us in a way that taken multiple input variables which is simply more than a volume-weighted shrinkage factor.

So, I think that, you know, so I think that we're going to have to ultimately think about kind of where do you draw the line on these types of measures which are fundamentally apples and oranges with the all-or-none in the, you know, several different other flavors.

Karen Pace: Right. OK.

Lyn Paget: Karen?

Karen Pace: Yes.

Lyn Paget: Can I – this is Lyn Paget. Just a quick ...

Karen Pace: Sure.

Lyn Paget: ... historical question. It sounds to me and as I read through what you (said) that when the committee did its work back in '08, '08 that there must have been a rationale to keep the definition as vague as it is, and I just didn't know if there was anything that would be helpful to know at this point because from what I hear, it seems that there could be a reason or justification to be thinking about a different classification system or at least a more substantive definition and I know that that's one of your objectives, but I just want to know. It seems it was – it was defined or it was created in a somewhat vague way to begin with and was that, you know, I assume was intentional was that to had not prohibit certain measures from coming in with, you know, what was – what was the thinking behind that.

Karen Pace: Right. I'll say something and Dave Shahian may you want to and Steve Wright may want to add to it to see how their memories compare to mine. But I think at the time, you know, we had identified there was no kind of standard taxonomy or classification of composite measures and there was a recognition that there were multiple methodologies for combining measures, and I think it was intentionally broad, and I think, you know, it was kind of where we were at, at that time with composite measures for quality performance, and I think we're identifying that broadness has created some or has left some confusion in terms of then what criteria need to be applied.

So, I think it was more of reflection of, you know, where we are in the field at that time but I'll see if Dave or Steve or anyone else wants to comment on that.

David Shahian: Dave – actually was sheer cowardice ...

Male: I would agree with that.

David Shahian: No, I think we intentionally kept it broad; it was the first NQF venture into this area. And I think we didn't want to exclude any types of composite measures that might not have thought of, and I think we anticipated that there would be the kinds of questions that ever risen such as the ones that Karen has talked about, and I think this is a perfect to (flush) out those vary broad guidelines we developed at that time.

Karen Pace: OK. We can always come back to this, but I think I'll move on to the next thing in the (briefing) memo and on the slide that we've identified and that part of the guidance and criteria that were developed in that early work had to do with the individual components need to be NQF endorsed or evaluated as meeting the NQF criteria.

So, what we found is that it's very difficult to implement that. It's unclear what is required when a component is not NQF endorsed and this idea that, you know, a measure might be important enough on its own to get NQF endorsement but could be important in composite. It's really hard to get handle on that because if it really met all of the NQF criteria then why wouldn't it be suitable for endorsement as an individual performance measure.

So, we really need to look at that criterion about the component measure should be NQF endorsed or be able to meet the criteria. We've had the issue – some having interpreted that language of not important enough meaning that it doesn't have to meet our criteria under importance which are high-impact opportunity for improvement and evidence. And it's kind of hard to think about why you would have a component of a composite that didn't meet those criteria.

So, another question that's related to that is what people sometimes called "balancing measure." These or, you know, that's probably jargon but it's often used in Q.I. that if your goal is really a particular performance measure but you want to have another measure to kind of tracking potential untoward consequences of performance measurement then, you know, those won't be endorsed as individual performance measures but do they have any potential role in a composite.

So, for example, if you're really trying to focusing on hospital length of stay you may have, you know, that may be happening at the expense of patients being readmitted very quickly. So, that would be an example of, you know, the performance measure might be length of stay but you need – people might consider a balancing measure of looking what impact that's having on readmissions.

So, that kind of falls broadly into that category, and again how to apply this to all-or-none measures when they're not really combining individual performance measures in terms of a totally separate individual measure but certainly the numerator components could have been in individual measures. So, it – that's really that language and criteria about they should be NQF endorsed. I'll say that the intent of that was to really, you know, encourage the combination of NQF-endorsed measures to really be able to focus on the composite then in terms of the methodology that we've – first of all, the conceptual construct and the methodology to combining those rather than having to look at all those basic criteria.

So, I think the intent was to really be able to focus composite evaluation on those things unique to it being a composite. But the reality has been a lot of times composite that are submitted with components that are not already NQF endorsed, you know, how to apply that to all-or-none measures et cetera.

So, I'll again stop there and see if you have any thoughts or clarifications or anything to add to that.

Alan Zaslavsky: This is Alan. Sorry, I didn't identify myself before. I think Sherrie articulated one of the reasons why you might create a composite when the individual measures are not considered important which is that the individual measures might have very low reliability or what not a little bit more. They might pick up only one component of a domain and be too partial but as you put it bunch of them together then you get something that actually in combination becomes important and (interestingly) be viable, to be useful. So ...

Karen Pace: OK.

Alan Zaslavsky: ... that seems like a reason, you know, just a general comments on the last discussion is that part of what's going on here that there are several different rationales under which something that (broadly) might be called a composite could be created. This is one of them and there are others and the guidance they try to have only one set of criteria when there are multiple rationales might not really be adequate to deal with all cases.

Karen Pace: OK. Thank you. Yes ...

(Patrick Romano): I think that's a very good point Alan, and it reminds us that perhaps we should have an approach where measure developers would enunciate what's their rationale is for composite. And then the type of information that they provide would follow from that enunciation.

So, if the purpose of the composite is to have a measure with reliability as Sherrie has already articulated then obviously the focus is going to be on the extent to which compositing the measures achieves that goal of enhancing reliability. On the other hand, there may be other rationales and both will

have to kind of work through this logic a little bit more, in which case, the focus of discussion in the steering committee should be different.

Sherrie Kaplan: Yes, this is Sherrie. Just to underscore what both Alan and Patrick just said, I think that if that was Patrick – they – the idea that you're trying to represent something, some abstraction like in the case of combining length of stay with readmission to try to represent quality of sort of general quality of performance of the hospital combining mortality with other things. You're also trying to represent some abstract construct and then the question becomes, you know, how many things do you need to look at to correctly represent that construct.

So, I think the purpose of measurement as Patrick and Alan referred to should stir that first and that should be the first thing that needs to get represented correctly. What are you trying to measure and why?

Dale Bratzler: This is Dale. The other thing that's come up with composite measures are at least ones that we worked with is that, you know, it's actually fairly easy to create a composite that looks at the group of let's say processes of care and you're looking at and you can calculate the score and evaluate compliance with the group of processes or an overall score. But when you look at whether the composite is related to patient outcome that's when you try to do that linkage then the individual components of the composite become very, very important, and we've actually seen some composites that predict worst outcomes based on some of the components that are in the composite and it's not that the component was bad but the way the component affects the composite itself adversely affects the linkage between process and outcome.

Male: Right.

Karen Pace: OK.

Male: They need to build on that if you (were) a measure developer to clear that you're aim in constructing a process composite is to represent the aspect of the process of care that improve the patient's outcome with mortality representing that outcome, then it becomes very important to demonstrate that each of the components actually adds information related to the prediction of

mortality. On the other hand, if your conceptual framework is not wedded to a connection to outcomes or the mortality then demonstrating those associations might not be as important.

Karen Pace: OK. Good. So, I'm going to move on to the next one and just so we can come back to any of these but the next issue which is kind of gets to the question of, you know, what are the implications is what we've set-up is testing requirements for composite measures, and we – the prior guidance was that, you know, the composite measure needed to be – determined to be reliable and valid. Obviously, we've already talked about the individual components should meet criteria which means they would be reliable and valid, and I think the point about an individual measure might not be that reliable on its own and I think that's something important that we need to consider.

But we also identified additional testing requirements for composite measures. So, we wanted component analysis to demonstrate that the included components fit the conceptual construct. Analysis to demonstrate that the components actually contribute to the variation and the overall composite score, analysis to support selected weighting methodology and analysis of missing component scores.

The entire guidance, you know, address that not all measures – composite measures may be based on psychometric approach that sometimes composite measures may start with just you have a group of measures than you want to come up with an overall score. Now whether – and that may be a reasonable thing to think of with composite measures with may end up deciding that it's not.

But the idea was that, you know, a lot of these, you know, are pretty much grounded and psychometric approaches and should there be different testing, you know, and I think the question that's already been brought up is do this really applied all the different reasons you might and things that you might call a composite, and if there are differences that we need to further clarify that so that we have it right.

And I think, you know, we do ask in the – I think that was also one of the criteria I didn't list to here but it's in your documents. You know, we do – one of the initial things that we ask for and that needs to be provided is the quality construct. We also say that, you know, if a composite measure doesn't really fit the psychometric approach to analysis that they should provide other analysis and describe what they're providing and why.

I'll give you one example of, you know, for example, we may have had a composite measure submitted where they did more psychometric analysis, factor analysis, inter-item correlations et cetera which really turned out pretty poor but they went ahead and submitted that to us for evaluation. You know, without considering whether that was really even the appropriate analysis or any kind of consideration of why those didn't really support the composite. So, we just, you know, seem to have a need to really nail down this what is required additional testing beyond our standard criteria, and if there are differences about, you know, these different types of composites or what will end up discussing our different types of composites.

So, I'll stop there and, again, if you want to add anything or question anything.

John Birkmeyer: So, this is John. Just a point of emphasis probably not adding too much that hasn't been put on the table in various forms and that goes back to I think that one of the very important early test of this reevaluation process will be to consider a taxonomy or some labels of kind of the small handfuls of different types of composites because those criteria for evaluating ...

Karen Pace: Yes.

John Birkmeyer: ... on their scientific rigor, you know, are completely influenced by whether their composite measures of, you know, of standalone processes of care or whether there are any outcome measures, whether there are sort of accountability and (steerage) intended measures or whether their compliance and Q.I.-oriented measures. So, I, you know, so I think that there's some real opportunity to, you know, add some clarity to the field there.

Karen Pace: OK. Thank you. Any other?

James Chase: This is – this is Jim Chase, too. In this area, I don't know if this will be relevant to our work, but I think there may be an opportunity to divide some of these composites related to the testing. If there are different for example – different – sometimes I think there are different denominator populations or different samples and ...

Karen Pace: Yes.

James Chase: ... a measure submitter may be pretending like that it's not there – by the way, they construct the measure but indeed there is sort of a subsample going on if you excluded certain patients for certain components.

So, I don't know if we'll have a role in providing some advice about how to deal with those situations, but it does seem important if the different level of rigor we may need with some of these composites versus others that have standardized denominators.

Karen Pace: No, I think that's a good point, and certainly worth consideration and I think, you know, the other point that was made these really are evaluation criteria for NQF, but obviously, they'll have implications for those developing measurement that are, you know, that are planning to come to NQF for potential endorsement. So – but you're right that it's specifically for how we would evaluate measures then what kinds of data they would need to present.

The other thing to just keep in mind kind of an overarching context that we need to keep in mind is that NQF endorses measures that are intended to for use both for performance improvement and accountability application. So, if, you know, we don't endorse measures that where the measure developer would say this is only potentially useful for quality improvement. We don't think it has any role in accountability that's generally we would not be endorsing or considering that type of measure for endorsement.

OK.

Patrick Romano: Patrick again.

Karen Pace: Yes.

Patrick Romano: Just a comment kind of tying together of things that we talked about it the very interesting issue that Jim raised – has developed a denominator because, of course, we have a common situation with outcome measures where measures have poor reliability because of their rare outcomes, and the traditional approach is problem that it goes back to many, many years of clinical trials is to create a composite outcome where there's a simple Boolean or logic that you've either had event A or event B or event C or event D and that becomes a single measure with denominator that's of course shared across all those outcome – across all those outcomes. So, we're used to thinking of that as being a single measure with a composite outcome.

Karen Pace: Right.

Patrick Romano: But then if you take the same individual measures and you estimate each measure separately and then average the scores then it's a composite. So, that a tricky distinction and, you know, we'll have to be – we'll have to think about both the semantics of that, as well as the (methodologic) issues.

Karen Pace: Now, good point. Thank you.

OK. The next issue that we identified is just to need to discuss again the issue of combination of process and outcomes measures into a composite measure. We have some that do that. It's a question that comes up of whether that's a good approach or whether for example if you have multiple processes or steps of a process that need to be followed, would it be a better approach to have a composite of those processes paired with the ultimate outcome measure or – and, you know, or whether it's (fair) game, you know, depending on the developer's aim and construct but doesn't have any implications for our evaluation whether it's all process, all outcome or a combination of process and outcome.

So, I'll stop there.

OK, and then the next one, and we can come to any of these but, again, it gets that kind of what goes into a composite and whether it has any implications

for evaluation or for taxonomy as we've been talking about and that's crosscutting versus condition specific composite measures. I think, you know, sometimes, you know, one of the at least one of the drives for composite that was discussed in the prior project was the interest in combining multiple measures for ease of use whether it's for a consumer trying to make sense of, you know, (six) measures in a particular condition or whether it's for looking at rewards and incentive, but then they kind of – when it's – I think the discussion also goes sometimes with the composites that are submitted that when it crosses multiple conditions, say, outcome of a lot different surgical procedures.

The question is how useful is that, for example, to a consumer if they're actually going to that hospital for one of those procedures and how transparent is it to be able to find out what's in that composite and use – the drill-down information, however, as we've already talked about if you get down to the – if composite was really constructed to increase reliability then, you know, the usefulness of the drill-down information if that's not so reliable is questionable. So, I think it's just again an area for us to have further discussion and clarity about, you know, those kinds of measures and how we need to look at them or if there are any differences in how we look at those kinds of crosscutting composites.

Dale Bratzler: So, this is Dale. I'm actually at the CDC HICPAC meeting and this topic actually came up for a discussion this morning as CDC starts to think about building the infection composites. And I think part of the leadership made the compelling argument that there are different audiences (for) the measures acknowledging that you lose some of the major specific detail in the composite but sometimes particularly for policymakers or others rolling up measures on the composite is useful.

So, we discussed all those issues this morning, and I think they're still – was at least in the group strong support for, you know, composites for some purposes.

Patrick Romano: This is Patrick. I think the challenge we have here is that NQF is forced to make this dichotomous decision of endorsement or not endorsement and that decision really over simplifies the world, right, because ...

Karen Pace: Yes.

Patrick Romano: ... viability and validity are relative concepts and there is no absolute threshold of what's reliable enough, what's valid enough. So, how do we deal with that? I mean, we have to same problem, of course, in clinical decision making where we know as clinicians that we work in a probabilistic world, and there are very few patients where we're 100 percent certain of the diagnosis or the right course of treatment. But we have to make the best decision we can given limited information and given the best effort to estimate what the probabilities are for that patient.

So, we're confronted with that here to that, you know, it's hard to say that the drill down is the sufficiently reliable because it's really a question of not sufficiently reliable according to whom, for what purpose, and those are complex regimens because the concepts are relative not absolute.

Male: I think another aspect of this issue that Patrick raised of things we had endorsed dichotomously by the purpose – (whether) the purpose of that – the reasons for doing the composite can really vary quite a lot and, you know, one of kind relation is the kind of integrated composites rather than as opposed to single construct composites. You might or – and might want to come out with a composite that overall have to a quality. Summarize, of course, everything and we know that overall hospital quality isn't really one saying and yet someone, you know, there might be a purpose for which that be useful thing and it could be done better or worse.

Would we say we never do that because it wouldn't (serve) the standards of psychometric (validity) that we look for one when we – when they created a composite of survey items that are supposed to be giving you reliable idea on one topic. I don't know that that's necessarily the case but the purposes are the kind of an integrated composite that really meant to summarize a number of different things into one measure is really different from the criteria or

something those purposes to be a measure of – really a single construct that you improve the reliability for by getting more items.

Nancy Dunton: This is Nancy. I think that one of the issues related to this is clearly the rationale for or the reason for creating the composite and things that cross measures – crosses the cross measures are potentially useful for consumers who want to pick out a hospital that's covered, you know, pick out an insurance plan out of their options that cover different hospitals or they already have a particular need, specific kind of care, they may be interested in composites related to that particular outcome. So, it seems to me that composites are less useful for internal quality improvements unless they're – the all-or-none composites around that process of care related to specific outcome than they are for public accountability.

Sherrie Kaplan. Yes, this is Sherrie. Back to something that Alan just said, you know, if an individual item that you're using multiple individual items each of which is so reliable that it kind of – it's useful and precise on its own. It defeats – it's actually counterproductive to add those kinds of things up. You'd end up over sampling the things you need to represent whatever abstract construct you're trying to represent. So, actually there's a kind of a tension inherent in the creation of a composite as the notion that I use one single thing to measure this complex construct to get it right.

You actually need more things to represent a construct – complex construct for exactly that reason. So, using – so if NQF is truly trying to do – it has to be useful for both all of the elements of a composite have to be useful both for assessment and improvement than you've kind of got to a really – you've got to a paradox going on. So, how well are you guys to having that be true for complex consequently, you know, for composite measures?

Karen Pace: Well, I think that's a good question, and I think that – I mean that's just a general overarching principle of NQF endorsement if there are really, you know, strong rationale that we need to somehow view that differently with composite measures, we need to, you know, carefully layout that argument and, you know, these – what come out of these will go to public comment, and ultimately, to our CSAC board.

So, I think, you know, that's, you know, we want to get this group's best advice and thinking so that we can, you know, if we need to question some of those assumptions, we can certainly do that. We don't want to cut that off from the beginning, but we, you know, we'll have to carefully layout the pros and cons and rationale and get further broader comments and input on that.

OK, the last thing that I'll mention that was on our list was measure harmonization and we won't go into – we don't need to talk about it in any great detail but it's just to note that when we start talking about composite measures that are a combination of individual measures that it kind of multiplies our issues with measures harmonization and which most of you know has been a challenge in and of itself when we're even talking about individual measures.

So, you know, I think the other things we talked about need to be addressed first, but just something that we can at least keep in mind and we may end up with some suggestions about that.

So, I guess ...

James Chase: Karen?

Karen Pace: Yes. Go ahead.

James Chase: This is Jim. I'm sorry. I needed to backtrack for a second if I could.

Karen Pace: OK, sure.

James Chase: At the – I missed – I want to see if you could help explain to me in the report that you sent out for this meeting right above the testing requirements for composite measures that last bullet point. I wasn't real clear on them that the last bullet point under the individual components where you talked about where the components are aggregated at a patient level and not necessarily based on the individual performance measures. Can you tell me more about what the issue is there?

Karen Pace: Are you talking about the briefing memo? I'm sorry.

James Chase: Yes.

Karen Pace: What page is that on? I just want to ...

James Chase: I'm sorry. It's on page five.

Karen Pace: OK.

James Chase: The fourth bullet point down.

Karen Pace: Analysis of missing component scores? Oh, I see, above that you mean?

James Chase: Yes.

Karen Pace: Yes, I'm sorry. OK, about how to implement the guidance for all-or-none measures?

James Chase: Yes.

Karen Pace: OK. That's in relationship to that idea of the individual components would need to be NQF endorsed. So, for example, if you want to take the all-or-none diabetes measure for example, we may or may not have individual measures that address each of those components of the diabetes composite, and so if we don't, well, there's a couple of issues there. One is if we do, we have – we tend to not see that that the all-or-none composite is exactly using those prior endorsed measures even if it's just the numerator component.

The other thing is, you know, how do we apply the evidence criterion to each of those components. So, I – it's probably a little unclear but the – it really relates to that whole idea of – if we're saying that the individual components of a composite, and we haven't made a distinction on these various types of composites if they need to be endorsed how do we actually apply that to an all-or-none where it's not really just combining the scores of the individual measures.

James Chase: Thank you. I was confused. I thought that you were trying some point about the individual patient performance (insight).

Karen Pace: Oh, OK.

James Chase: That's not – not what with (many) – individual – the individual components of the measure were going to ...

Karen Pace: Right, right, right.

OK. So, one of the things we wanted to just get your advice on is whether you've identified additional issues and then we'd talk about this in different context and some additional discussion about them but is there anything that wasn't touched on in this memo or that we've already discussed that we need to make sure it's on the table for exploration with you all?

David Shahian: Karen, it's Dave Shahian. You – I had to step away for a second. You may have discussed this but ...

Karen Pace: OK.

David Shahian: ... I recall one of the things during our initial steering committee discussions that – I'm not sure we ever really resolve was how composite – proposed composite measures were going to be evaluated by NQF. In other words if their surgical measure that is going through as a number of other surgical measures. There are additional technical complexities to evaluating a composite measure. So, should a measure that is going through another process be separately evaluated by either this entire composite panel or representatives taken from this composite expert panel so that – so that the necessary technical expertise has really been available to evaluate the composite in addition to the usual sorts of evaluation we have.

Karen Pace: Right. Good question and as you know the first project we dealt with the evaluation framework issues and then that group actually evaluated the first set of composite measures. Since that time the measures are submitted in the topic area and that is also a source of something we have to deal with in terms of the expertise, in terms of evaluating some of these additional testing requirements for composite measures and how best to address that.

So, I think, you know, that certainly is something that we can talk about and see what the recommendation is. It does become problematic. Our steering committee members are all volunteers, and they only have so much time to try to learn new things if we're – if, you know, for example how to look at a composite measure. So, I think that's a valid thing that we need to have some further discussion about as well.

Sherrie Kaplan: This is Sherrie. I wondered if there is any – if you – if this is too complex to tackle but the attribution target, you know, whether it's – whether the measure is supposed to be attributed to the physician, the clinic, the institution, you know, where are the components of variation for and how we suggest people tackle the problem of evaluating at what level if the precision reached sufficient, you know, reliability or (just) come validity may escape us, but for the purposes of evaluating and improving quality where the attribution target is, and how confident are we that the measure is appropriate – the composite measure is appropriate for level of attribution.

So, if there's too much variation at the patient level, do you add up patients per doctor? And if so, how is that evaluating? You know, what are the sampling strategies to do that to get a reliable physician level score and then how many physicians do you need to get a clinic level score and how many clinics do you need to get up the institution level score. The only reason I'm raising is this and Alan actually is involve in this too.

There's sort of now – an interesting how that sampling strategy goes on and what are the precision issues when you're to create a composite representation of quality at different levels of the health care system.

Karen Pace: Good. Yes. I think that is something we should explore further. I think that discussion also has implications for our – say individual performance measures that often comes up in terms of the level of analysis and at what level, you know, has been tested and demonstrated that it can have precision or reliability for that particular level, but I can see that that might be the more complex for the composites and I think we should address that.

Male: So, seeing up the agenda for our in-person meeting.

Karen Pace: Yes.

Male: I wonder if – I'm getting kind of sense that – I'm not sure that we need to tee-up measure harmonization as a separate topic.

Karen Pace: Right.

Male: Because I think that, you know, that issue has been discussed elsewhere within NQF ...

Karen Pace: Yes.

Male: ... for any unique issues as you say, they may be magnified for composites but I think there are other mechanisms for addressing that issue. But what I am hearing is definitely (center) in discussing the NQF processes around composite evaluation, and how those processes can be strengthened to provide both the content area of expertise that resides in the steering committee and the methodologic expertise on the (resides) and expert panel like this one.

That's going to be the real challenge going forward because I think many of us who brought measures to steering committees have had the experience that, you know, they're very good at understanding the clinical specification issues. They're not so good when you try to explain, you know, a complex scoring scheme that incorporates a different constructs in weighting composite measures.

Karen Pace: Thanks. Good.

OK, any other key issues or questions that we should be sure that gets on the agenda for the in-person meeting?

Lyn Paget: Karen, this is Lyn Paget.

Karen Pace: Yes.

Lyn Paget: And I'm not sure I'm going to piece this together logically or not, but a couple of the themes that I heard one kind of pushing back on the protocol that NQF have around what you need to define your measure to be in order for to be

submitted for endorsement and that whole only quality improvement question. I think that that becomes fairly significant.

The way I understand the purpose of these measures is it is simplification exercise. It's to be able to kind of pull together meaningful numbers and present them in a simplified way, but I think it also calls into question and instrument that may be too blunt – a tool that's too blunt that it's only advantageous for pushing broad shift.

When I think that, as I understand, measurement trends with the support of I.T. and electronic health records, we should be able to take a much more new ones look particularly at patient experience and I wonder if there might be a way to kind of look at this from the lens of trends and to say – or to spend a little bit of time around these particular composite measures and the role in measuring to advance the kinds of changes and culture and clinician behavior and so – and so forth that we're aiming towards as we move towards different payment structures and population health management and so forth.

I mean to me this feels like an opportunity in that whole simplification aim to bring those kinds of measures forward and not restrict the NQF in a way or restrict the recipients of the endorsed measures. So, the whole role of these in that context seems really important to me and then I also just wanted to throw out for – because of the curiosity that I have for the measurement experts. The trend for colleges to make these standardized test scores such as the (SAT) optional.

I understand is tied to questions around the predictive ability in those standardized tests and I know that that's modeled as a kind of basis for these measures, and I just wonder again, you know, where the trends are going in education versus health care.

Karen Pace: I don't know if anyone on our panel has crossed over in the education, but certainly, a lot of the psychometrics have a root in educational measures but I don't know if anyone wants to comment on that or ...

Sherrie Kaplan: Actually – this is Sherrie. Sherrie, (Elisa) and I – Sherrie, (Elisa), (Norman) and I looked at education as an example of where standardized testing is used

to estimate teacher performance and that caused a whole lot of controversy on how much of the (variance) in children's scores is attributable to children in the family versus the actual future performance et cetera, et cetera and then you winded to school district and so on.

So, there's a whole host of things going on in education right now that are related actually to what's going on right now, and we can have a chat about that I think in the in-person meeting if anybody is interested.

James Chase: This is Jim. I am interested though and I think something you touched on there related to is ease of use a, you know, part of the criteria? Because I worry what composite – some of the reasons why we're doing them obviously is that and not just for the public. We now that works for the public but we found I think even with clinicians having some focus around some things may be important and the, you know, we may always be able to find that for many composites, the individual components will be more reliable and incredible and yet there is a reason sometimes to want the roll things up just because it makes them more able to be used.

And so I'm not – I'm curious about the thoughts that NQF about – is that a valid criteria? I haven't seen it anywhere but I would one to argue for it should be able to be addressed. Obviously we don't want to make ease of use be (trumped) everything else, but it should be a criteria that can offset some other things.

Karen Pace: Right. And we've just recently kind of reframed our criterion about usability and use but certainly feasibility is, you know, something to consider and I think what you touched on is one of the reasons people actually started looking at composite measures than the quality performance arena because it may be difficult especially if you're considering the consumer audience while – whether you're a consumer or a professional audience if you go and look at six performance measures and you see, you know, if they're all in the same direction that may be useful, but when they start maybe being in different directions, what do you make of it.

And I think, you know, that is certainly one of the things that people certainly talk about in terms of a rationale for the composite measure, and yes, how we might want to move that into criteria or how we look at the rationale that we've been talking about, you know, we can certainly address.

Patrick Romano: Yes. And this is Patrick. It ties also to the (Brent) policy interest in value-based purchasing or pay-for performance schemes.

Karen Pace: Right.

Patrick Romano: At the end of the day, almost all of these schemes are composites, right, and they're always (of) taking different measures and translated them into a single unit of dollars and then the hospitals or plans with better performance to get more dollars and that is the ultimate signal is the dollars. So, if those – if that final product – the grade, the dollars is being reported separately then by definition it's a composite because it's coming from a variety of different variables and, of course, it's not necessary that those variables are related to each other because the hospital may get a reward for doing well in A but not so well in B or by vice versa.

Karen Pace: Right.

Patrick Romano: The highest reward for doing well in both. So, there's a (policy context) with this discussion related to pay-for performance in particular.

Karen Pace: Good.

Male: I wonder if that would help us in our in-person meeting if someone get would start to develop, you know, this – table of some of these different kinds of composites. The notes on this and the briefing memo is to be more focused around things like all-or-none versus none-or-all or whatever that they're detailed about how they're computed, you know, sort of more, you know, procedural kinds of assets to them, but I think we've been talking more in terms of the rationales, in terms of why they – why they're put together the way they're put together for the composite and what is being used for.

Karen Pace: Right. I think that's a good point. You know, when the original group met, we seem to look at more about the differences in methodologies, but I think it's almost sounding like – I don't know if it's more important or just as important with what's the original reason for combining these things, and I think that's a good point that we could at least have something drafted and, you know, for us to explore further at the in-person meeting.

OK. So, we can always come back to this but again some of the other things we wanted to just see if you had any input on is any key references we shared with you the draft of the prior composite report plus the white paper that Sherrie Kaplan was involved with that was kind of background work for that group and also we just I think sent out or posted today a list of references, and I know many of you may not have had a chance to look at that, but we would certainly be interested in – if there are key references that you think we should make sure that we have for this group's consideration as we're looking at specifically, you know, evaluation criteria or things that will help inform your thinking and the work of this group.

So, you know, you may not be able to do that right at the top of your head, but we would certainly be interested in that if you've identified anything that we should be sure to distribute to everyone.

And then really the last key question is really related to both of those is if you have any suggestions about the approach or resources for the in-person meeting that we should think about in terms of planning that meeting. We're definitely interested in your suggestions, and I guess maybe I should help or talk a little bit more of what we want the output of this group to be.

So, I think the key thing is that we want to develop guidance for evaluating composite measures that are submitted to NQF for potential endorsement and, you know, I think the whole reason we're talking about this taxonomy and the different reasons for creating composites and different methodologies is that it definitely relates to evaluation criteria whether there are differences then based on those things of what testing needs to be done to demonstrate that it's suitable for endorsement, but I think ultimately that's our goal is to be able to develop guidance for NQF in evaluating this composite measures and, you

know, our typical process is we would develop a draft report, certainly the thinking and rationale of the group but ultimately, you know, what are those criteria that are being recommended that, you know, these composite measures should be evaluated again.

And so, I just wanted to put that in context of what our ultimate product or goal for this group is but certainly interested in any of your suggestions so that we get this meeting planned, you know, so that it'll be most effective and efficient for discussing and making recommendations.

Male: One thought I had is when we get sort of to a draft level of some of those recommendations or guidance is to actually test them against take a little time as a group to test them against some existing measures, but maybe more importantly is do a little brainstorming around, you know, what's around the corner, what's going to start coming from our opinion, and how will those layup against this guidance. Just to make sure we haven't missed something or we're setting up some kind of road block that we didn't intend.

Karen Pace: That's good, and I should mention that one of the things that we're doing, I mean, we started by just giving you a list of endorsed composite measures, and we've already talked about some of our challenges in terms of our database and also the confusion about what is the composite but we were going to actually look at those and kind of look at the methodologies and some of the key things that we'd talk about, but I think, you know, so we can certainly, you know, have some of that more detailed information that we may use to look at this against. I think that's a good idea.

Karen Johnson: Does anybody have – I'm just curious in terms of organization for the meeting. Do you like breakouts? Do you think that would be useful or not useful? Do you like doing those brainstorming sessions where people talk Post-it notes and stick on walls? Or do you like things that are more like things that are more like our usual steering committee meetings where we sit around the table and talk? Anybody have any ideas about that?

Male: It probably depends probably the terms of what you're – exactly what you're hoping to have going out of the meeting in terms of conclusion, you know, if

the staff is going to figure out from a very, you know, unstructured conversation. What our recommendations are, that's one thing, or if we're trying to come up with some pretty specific recommendations I feel that somewhat different process.

Karen Pace: Right. Well, that's a good point. So, we will definitely give that some thought of we did – that's one of the reasons we purposely already set up a call to follow the meeting is that we expect that we'll have to be doing some drafting and verification after the meeting as well, but I think that's a good point, and we can think about that in terms of the time. I mean, this expert panel is small enough that we can all get around one table and have discussions so that will help to a certain extent.

So, we'll definitely keep that in mind in terms of, you know, structuring the meeting, but again, on any of these things if something occurs to you following this call, please send us an e-mail because we're going to be, you know, very much involved in getting ready for that in-person meeting.

Karen Johnson: One of the things that is a little new to us, we do have what we called discussion capability on SharePoint where you can start a thread and then people can reply in that sort of thing. So, we were thinking we might just go ahead and seed a few discussion topics maybe the ones that Karen has already addressed and maybe the ones that you have added to it, if you think that would be of use and that you would have time to think about before the call, you know.

Male: I think that would be helpful. It maybe, as you said, since we have some subject areas that we've come up well sometimes breakout groups help to get through the work a little faster if you can divide it up a little bit and then bring it back to the group of whole with some recommendations people can react to as opposed to the whole group having to go through each subject area together. But again, that may depend on sort the time that we have in the meeting and how that shakes out with the different areas. So, I would leave that to the chairs.

I'm not as in favor of breakout groups on the same subject and then coming back together. With the group as small as this, I don't think that's really necessary ...

Karen Pace: Right. OK.

Male: A lot of – just depends on the preparation and then I think the more we can have some of the issues structured and so – and the things we can look at going into the meeting and then what we are going to get.

Karen Pace: OK. All right. We'll keep – if again, you know, feel free to just send us an e-mail and we'll be keeping you updated on SharePoint. We'll let you know if we start any of those discussion threads. We may just set them up and then we'll see, you know, how that goes.

And maybe what we should do is see if there are any public comments from the public that may be on the call. Operator, if you want to see if anyone has any comments or questions that they want to bring to our attention?

Operator: As a reminder, in order to ask a question, please press star one on your telephone keypad.

Once again to ask a question or make a comment, please press star one.

There are no questions.

Karen Pace: OK, I'm going to turn it back to Elisa to talk about wrap-up and then, you know, we'll have one last chance if anyone has any comments or questions.

Elisa?

Elisa Munthali: Thank you, Karen and Karen. So, as both Karens mentioned, we are going to continue to gather literature on composites related methodology and any changes that have happened since our last project, and we're really appealing to the panel to pass along any references that you may have, it will be definitely helpful for us to include on the list – the preliminary list that you received yesterday.

At the same time, we're going to continue to go through our database to analyze the submissions that we've received so far in composite, and with regard to criteria that I'm going to share with you later and so that we'll have a very comprehensive list for you for discussion. And all of this including today's discussion is in preparation, again, for our in-person meeting on November 2nd – it's Friday, November 2nd.

And I just wanted to make sure that all of you have heard from our meeting's team to plan your travel and accommodation here to Washington, D.C., and if you haven't, please get in touch with me. Our e-mail address is on the screen both Karen Johnson, Karen Pace and myself and our telephone number to the main line is there, but I'm sure I've communicated with all of you by e-mail so you have my direct dial.

So, are there any other questions with regard to logistics or with regard the upcoming meeting?

Operator: Once again, to ask a question or make a comment, please press star one.

Karen Pace: And this Karen Pace, and we'll also schedule a time with the co-chairs prior to the meetings to kind of again to help solidify our plans for that meeting.

Female: Well ...

Operator: There ...

Female: Sorry, (Raquel).

Operator: I'm sorry. I'm just going to let you know, there are no questions.

Elisa Munthali: OK, thank you so much, and thank you for joining us everyone, and we look forward to meeting you in just a few weeks actually, and you'll definitely be in touch with us. We wanted to remind you as well that we're doing – most of our document sharing on SharePoint. So, if you have any problems logging on, accessing the files, please do get in touch with us so that we can get our technical support staff in touch with you.

So, thank you very much.

Operator: Thank you, ladies and gentlemen, for participating in today's conference call.
You may now disconnect.

END