

f

NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
November 15, 2012
3:00 p m. ET

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

(Elisa Munthali): Hello everyone and welcome. My name is (Elisa Munthali), I'm a senior project manager with the National Quality Forum, and also with me on today's call are my colleagues, Karen Johnson and Karen Pace. And we just wanted to go over the objective of today.

So, this is a follow up on the technical expert panel meeting on November 2nd, and the purpose of the call is to review and modify the draft principles and evaluation criteria for the composite performance measures. We also want to identify any outstanding issues and make recommendations for the format and content of the draft report.

And so, before doing that, we wanted to do a roll call for a matter of public record. We've been taking roll of the technical expert panel members that are on the call, but I'll just go down the line, and in case anyone didn't speak up earlier, if you can just say you're present. So, John Birkmeyer?

John Birkmeyer: Present.

(Elisa Munthali): Dale Bratzler? Did I hear Dale?

Female: I think Dale's going to be late. He e-mailed us.

(Elisa Munthali): OK. Great. Jim Chase?

James Chase: I'm here.

(Elisa Munthali): OK. I think Elizabeth DeLong will be late as well, unless she's on the line.

Female: No.

(Elisa Munthali): Nancy Dunton?

Nancy Dunton: Present.

(Elisa Munthali): Liz Goldstein?

(Liz Goldstein): I'm here.

(Elisa Munthali): OK. Hi Liz.

Elizabeth Goldstein: Hi.

(Elisa Munthali): Sherrie Kaplan? OK. Lyn Paget?

Female: (Lyn's) not able to join us.

(Elisa Munthali): OK. Patrick Romano?

Patrick Romano: Here.

(Elisa Munthali): Hi (Patrick). (Dave Shahian)?

(David Shahian): Here.

(Elisa Munthali): Steven Wright? And Alan Zaslavsky.

Alan Zaslavsky: Present.

(Elisa Munthali): Hi Alan. OK. And I'll send it over to Karen Pace.

Karen Pace: Hi. Welcome...

Elizabeth DeLong: Hey, this is Liz DeLong, I'm also here.

Karen Pace: Oh, wonderful. OK.

Elizabeth DeLong: Sorry.

Karen Pace: No problem. So, thank you all for joining us and thanks again for your great discussions at the in-person meeting. Sorry, I wasn't there, but I've read every word in the transcript, so thank you.

So, I think you know, as Elisa already said, you know, one of the things we put in the – what we put in the briefing memo is what we think should be the mix of the draft report and obviously we'll have to put some context around that, but we really wanted to do a touch base with you about key principles, recommendations for the evaluation criteria, other recommendations and any other outstanding issues that you think we need to have some discussion about.

And with that, I guess I can turn it over to (Patrick) and (Liz) to see what your thoughts are in terms of proceeding through today's agenda and material and we can just dive in. So, Patrick, you want to make any comments or suggestions, and then we'll go to (Liz).

Patrick Romano: Well, you know, I have to say that (Liz) and I didn't really have a chance, unfortunately to confer with your staff about our approach to this, but I will say that I think that you folks did an excellent job summarizing the discussions that frankly, wandered a bit, but also covered a lot of terrain.

And I think that this is excellent start in terms of stealing some of the key points from that discussion. So, thank you very much for all the effort that your team put in to this. It's a good framework to start our discussion, and I assume we'll kind of walk through what you've –what you've sent us. I'm open to recommendations and suggestions from others.

Karen Pace: OK. Yes, we thought we could do that and I'll just see if (Liz) wants to make any suggestions or comments and then, maybe we'll – we can just with the principle and see if anybody wants to bring anything up for discussion.

(Liz): I just wanted to echo what (Patrick) said, the staff have done a marvelous job of putting together the comments and making them concise and clear from the in-person meeting, and I think we should just go ahead and get started.

Karen Pace: OK. Well, thank you. So, we thought that we wanted to capture some of the key principles that – and we see this as things that would be kind of foundational that would lead into the actual recommendations about evaluation criteria.

So again, this is our first attempt to distill a lot of rich conversation and you know, if there are certainly anything that you think should be added, deleted or modified, and I think, just in the interest of time, we want to identify those things, but not spend too much time wordsmithing because we can certainly get red line word – exact word suggestions from your active call that I'll just stop there and see if anyone wants to address any of those principles or if you really want to go through them point by point, but we can...

James Chase: This is – this is Jim Chase. I have one I would...

Karen Pace: OK.

James Chase: ...I would start with, and so it will get us kicked off.

Karen Pace: Sure.

James Chase: So, under the principles, the fourth and fifth down, I just wanted to make a comment about and see what others think. I very much agree with the fourth one, we talked about that a lot around the composite performance should provide some added value over the individual measures and there needs to be some justification of that.

But then, when I read it in context with number five, you know, desire to create one score for multiple performance measures is not sufficient. It worried me a little bit around that issue that I brought up around. There may be – it might be better to have a principle that's more in the positive spin of

that resident thing what it can't be, because I'm worried that that could become a barrier.

So, my thought was more about talking about it may be an effective – a more effective use by public care providers multiple related measures may be a reasonable purpose, but then, you still need to prove the validity of the construct to the measure. And let me test this with people to see why I'm worried about this.

I think it's reasonable to do a composite, for example, of a prevention measure where you would bring together a number of prevention type of measures that may be stand on their own, but for ease of uses of patient, I'd like to know, you know, how that particular provider does across all of the recommended preventive services for people like me.

And so, that to me in some ways is just an addition of a bunch of measures that stand independent. And I worry that number five might make people think, oh no, you have to prove that there's something additive about bringing them together other than the ability to sort of look at something as a package.

Karen Pace: So, let me just – this is Karen Pace, and I'll mention one thing and then see what else other people want to do. The example you gave about a prevention composite seems like it would have a quality construct versus, for example, taking 10 unrelated quality performance measures just because you have them and putting them together. So, I think the example you gave is actually would be a quality construct, but I'll see what others have to say about that.

Alan Zaslavsky: Hi. This is Alan. I was – I was a little bothered by that one as well.

Karen Pace: OK.

Alan Zaslavsky: The way I am (inaudible) with it, and I think the language might have to be changed to make this clear if people agree on it. It's that the desire to create a single score is not a sufficient amount of evidence that we had with endorsing something as a composite, but it might be probably good rationale for one thing to do it.

Karen Pace: I see.

Alan Zaslavsky: Right. And this suggests that it's not – that could be read or misread maybe.

Karen Pace: Right. OK.

Alan Zaslavsky: Maybe it's not a sufficient motivation, in fact, it's often the main motivation we see in a composite is they have more personal use presentation that people can power more easily, and that – I think the point you're trying to make is that you can't just declare that you have your things to look at and that's enough reason to do it. You have to then demonstrate that it has some reasonable properties in terms of the objective of these interpretable as a combination of like things and (inaudible).

Male: Right. And I think that there has to be a coherent quality construct again to get back. So to put a positive spin on, the idea was that, I think the developers would have to explain what that coherent quality construct is.

And putting together every possible measure, for example, might not meet that standard because in the case of the prevention composite that you're proposing, a developer would have to say, look there are lot of healthy – basically healthy patients out there who don't have any chronic diseases who are concerned about how well their health plan or how well their physician organization provide preventive services to keep them healthy.

And to me, that could be a coherent quality construct, but it links to again the idea of decision making that there is somewhere in the market place, there are decision makers such as consumers who have to choose a health plan and this might be relevant information because if they don't have any chronic disease, they might be very interested in know what's the best health plan to keep me healthy, to prevent – does that make sense to people?

(Nancy Dalton): Yes. This is (Nancy) and I think that, perhaps, it would be a helpful guidance to developers to understand that they, in this circumstance, need to describe how the examples from the domains in the concept. These are not including, you know, a one-day list.

Karen Pace: OK. So, I think we – I think the suggestion is to award this one to positive and I think that's a good one and the key concept is that you have to have a coherent quality construct, you know, even if you start with just a list of 50 measure set available, you want to choose those that fit your coherent quality construct.

So, you know, if anyone has some specific wording suggestions, we'll definitely be interested in those, but does that kind of summarize the feeling then?

Female: I guess I'm a little confused as to why this particular board is needed.

Karen Pace: OK. So, you're right, because I think it may be...

Female: It's covered in the previous, isn't it?

Male: Yes. So, somehow right. I think it might make more sense to combine a couple of those forward.

Female: OK.

Male: Describe this concept.

Female: All right.

Male: So, it's again, it's taken out of the negative framing...

Female: Right.

Male: ...to the positive framing of...

Female: Right.

Male: ...of what I'm...

Karen Pace: So, I guess, and maybe this is something we can do in the text. So, this is the question to you because in reading the transcript, at least from my point, it seemed that perhaps there is some interest in emphasizing that just taking, I think the reference was kept going back just taking 50 measures that happened

– that we happened to have performance measure and somehow combining them is not a sufficient quality construct.

And so, the question is, do you want to, you know, maybe that's not a principle, is that something we should talk about, kind of in the narrative that might go with this or do you want to stay away from emphasizing that kind of negative altogether.

James Chase: This is Jim again. Excuse me, this is Jim again. I was thinking that it would be more helpful to again, may be combining these two of the first one around. It's not enough, I mean, if they have to provide what's the added value over the individual measures, and then, we also need to provide why these measures should be combined together.

So, we showed that it's both of those things as opposed to the mentioning around. You can't combine a bunch of measures together because again I think that's where the – and I understand what we're trying to avoid is just taking random measures in taking them together. You have to give some justification...

Karen Pace: Right.

James Chase: ...and show some validity of why these hanging together as opposed to because we have them.

Karen Pace: Right.

James Chase: Because these are all the measures we have on your orthopedic surgery and we though they'd all fit nicely together.

Alan Zaslavsky: Yes. I think that what we're trying to do is to say in a positive sense, whether the criteria the same as measure is well constructed, and then you don't have to say, but this other thing is good enough.

Karen Pace: OK.

Alan Zaslavsky: And we give what the criteria are.

Karen Pace: All right.

Alan Zaslavsky: And that surgery (inaudible) the quality construct of the composite performance measure are essential for showing what components et cetera.

Karen Pace: Right.

Alan Zaslavsky: And let's talk about the individual components, but I think that making sure that that point about they're being a construct and justification for the composite in terms of a construct is way...

Karen Pace: OK.

Alan Zaslavsky: That's the positive point that we want.

Karen Pace: OK. So, we can just take out the negative altogether, that's perfectly fine. Any other of the bullet points under principles that you'd like to discuss.

(David Shahian): This is (Dave). (Inaudible) the very last bullet, composite performance measures are complex and attention to be given to parsimony and simplicity. I think the parsimony and simplicity as I understand it should be in the way the results are presented.

But I think inherently, the construction of composite measures is complex and I'm not sure that the distinction between the complexity of construction and parsimony and simplicity of presentation is I think they're perhaps not those distinction aren't made clearly enough here or perhaps I'm off based, but I think that...

Karen Pace: Well, I think and it's good (Dave), distinction that we need to explore if (inaudible) that there was interest in parsimony in terms of the number of component, and also simplicity in terms of the waiting, which is, I thought more than just a presentation, but definitely others, what your thoughts are.

John Birkmeyer: This is John Birkmeyer. I agree with David. Basic point that there needs to be a distinction between the process of creating the measures and then the output in how those data are summarized or presented to the end user. Certainly, idea on the whole point of composites is clarity and (inaudible)

leveled measures so the latter piece, you know, I think applies to this bullet as it's written.

But on the other hand, some composite measures, just by their very nature, you know, are complex, and you know, and I don't believe that we should be evaluating them on, you know, fully on how simple we're waiting is, we should be evaluating them on, you know, on some combination of the – of the transparency of construction, but you know, but also their measurement properties and how well they discriminate and other factors at the end of the day.

Patrick Romano: Yes. This is Patrick. I'd agree. I think that the – what's off a little bit here is that there's an implication that a simpler scheme is necessarily better. And I don't think that that's really where we were coming from. I mean, a simpler scheme might be as good, but we don't want to preclude, for example, a factor-analytic waiting scheme, if that's well justified.

But we do want people to explain and justify their choice and the choices that they make. And if those choices are complex, there's nothing wrong with that, but they do need to explain or justify that complexity. I think that's what we were getting at.

When you deviate from parsimony or simplicity, we just need to hear why, and that can be then a topic for discussion within the steering committee. But I agree completely with David and John that simple waiting schemes are sometimes suspect.

Karen Pace: Right. So, should we take this out altogether or, I guess I think your right. If our efforts – other things being equal, you would take simple over complex or parsimony over lots of measures, if other things being equal, but perhaps, we don't even need this bullet.

Patrick Romano: So, I think the principle that needs to be announced is that if you're going to include measures in a composite, they don't seem to add any information. Then, in the interest of parsimony, you need to explain why you're adding those measures. If you're choosing a particularly obtuse or complex waiting scheme, you need to explain that.

So, I think that would be the principle. It's just a justification for why you're including more measures that it would seem are necessary based on empirical analysis, and why you're including a waiting scheme that's more complex than would seem necessary.

Male: I think that, you know, we agreed that simplicity and parsimony are nice things, that they're distinctly secondary to the way that the liability set a criteria.

Karen Pace: Right.

Alan Zaslavsky: And I think that could be expressed exquisitely with what you said that parsimony and simplicity are objectives but secondary objectives or secondary criteria that implies that a few primary criteria probably is something more complicated and you can justify that, but if not, then you should go to the simpler version.

Patrick Romano: Yes. Exactly. Something along those line.

Karen Pace: OK. Good. All right. Other bullets under the principles that you'd like to discuss?

Patrick Romano: One other question.

Karen Pace: Yes.

Patrick Romano: So, there's a statement here, I haven't counted the bullets, but...

Karen Pace: OK.

Patrick Romano: ...but the...

Karen Pace: I should number them.

Patrick Romano: ...way down. The reliability of the composite performance measure is a greater interest in the reliability of individual component and relief for validity. This is true, but it has to be interpreted in the context that not all of

the component measures are separately endorsed, because of course, if all of the components have been separately endorsed, then they should all meet at least some minimum standards for reliability and validity.

So, I think you may just want to contextualize some of those bullets a little bit to explain that this is in the context of moving away from the previous guidance and so that every component have to be there endorsed or to meet the standard.

Karen Pace: OK.

Patrick Romano: So, this refers to a particular situation where the developer is arguing to include component that are not endorsed.

Karen Pace: Right. So, what you're saying is, if they were endorsed, they would have met the basic reliability and validity. OK.

Alan Zaslavsky: There's a bullet, let's see, one, two, three, fourth from the last, the validity of the composite performance measure is a bigger interest.

Karen Pace: Right.

Alan Zaslavsky: And you say the composite might not be a valid summary of the valid items, and the reverse can be true as well. The individual items by themselves might not be valid, but in composite, there might be.

Karen Pace: OK.

Alan Zaslavsky: And we talked about some examples of that in the meeting. So, I think maybe (inaudible) has to be presented there.

Karen Pace: OK.

James Chase: This is Jim again. I have one that I don't know if it subject – I don't know if it fits in principles or not, but I'm just curious where it might. We talked a little bit about the process of looking at all the components of composite. And just from the way committees will work this, is there an opportunity for when a – if something's presented and a component isn't going to be endorsed.

We said, you know, people might be able to now say, I would like to have the individual components approved plus the composite, but would you be able to set this up so that if their one piece of it is not accepted, you can still move forward with the rest in the composite if you know all the validity and so forth checks out and the measure steward is interested in that.

Just going through this process as one, we found it unfortunate that we had – we sort of couldn't get endorsement because of one particular component of the measure and then we have to – if I understand, sort of get back in line to discuss this later.

Karen Pace: Right.

James Chase: And I would...

Karen Pace: So, that a good – that's a good question and it's actually come up recently in our two-stage pilot. And I think, the direction we're moving and certainly, we'd be interested in these group's discussion, but if the, you know, the recommendation there's one problem component.

And there's, you know, agreement with the measure developer that they're willing to address that and you know, can you know, justify that the remaining composite would be – will meet criteria. It doesn't have to wait but you know, I think that would be a question that your group's recommendation in terms of process, and whether that's actually going to be workable in the timeframes of our usual projects.

James Chase: Well, we – I mean I have a conflict of interest here probably, but we had a sneaky way of dealing with that as the staff knows in the previous process with our composite because a couple of the component measures were, you know, deemed probably not to meet the standards for separate endorsement. And say, we just said, well, we'll give those particular components zero wait.

So, we called the composite the same thing. We just manipulated the wait, so that effectively, those components are out as a composite, but it was kind of a

clever trick. It would be better to have a standard process for addressing that trauma.

Karen Johnson: This is Karen Johnson. I'm just wondering, I think I recall Sherrie making the point on that discussion that potentially dropping one of the components from the composite could actually invalidate some of the reliability and/or validity testing. So, did I understand that correctly, and if so, where do we go from there?

James Chase: And I was hopeful that they were there, may be two in some cases. In others, you may have enough validity on the other components where you'd say, there might have to be an opportunity sort of reapproach after saying, let us just go back and re-work some things and we'll – but I haven't go through a full resubmission.

And again, I'm not sure how the process would work, but it does seem reasonable that since you want people – we're going to make this so that it's hard to add a lot of things. Some people may be shooting for a couple that they think might work. And if they don't, it seems a shame not to be able to continue on with the components that do meet our criteria.

Karen Johnson: Right. And I think that's – you know, technically, we would prefer to have these things addressed in the project, which is happening. And you know, if that's something that can be done in the timeframes that we're dealing with, then that would be ideal.

So, I guess it could be as, you know, often happens, you know, the steering committee make that recommendation and then the measure developer respond, yes, we can do that and we can redo this analysis and bring it to you or you know, some developers will say no.

I mean, so I think, you know, but I think what I'm hearing is that, if at all possible to at least provide the opportunity for the developer to respond in the current project.

Patrick Romano: Yes. And I wonder – kind of as a coach here, maybe this belongs in a separate category. Maybe not so much of principle as some implementation guidance...

Karen Pace: Right.

Patrick Romano: ...I'm not sure if there may be a separate discussion that we'll have at the end to provide some guidance to NQF about how to implement these changes in a very practical operational way.

Karen Pace: Yes. And one of the things we have at the end of the evaluation criteria was additional recommendations and we can certainly if there are other implementation ones we can, but I think that's – for now, that's where we would put it rather than under principles and we can – but I think that certainly a reasonable recommendation to accompany this.

Patrick Romano: Everybody else comfortable with that?

Male: Yes.

Karen Pace: OK. Anything else under the principles or additional principles that you'd like to discuss before we get into evaluation criteria?

Patrick Romano: I guess one thing and it may be just something that I was fixated on after having read (Jeff Gephart's) paper and some other work, but I think in terms of describing the coherent quality construct, it would be helpful to see that describes with a bit, you know, about the decision-making context.

I mean, we know that NQF endorses measures for public reporting, but the way that you would construct a composite for consumer decision making is actually different from how you would construct a composite for an employer's decision making or for a Medicaid program administrator's decision making.

So, is that – is that out of line to suggest that the developers should actually describe how their measure should be used, how they envision their measure being used for decision making purposes?

Karen Johnson: I actually felt we discussed that, and I agree that they should, but I'm not sure I see a big distinction. We should all be looking for the same level of quality I would think. What am I going to look for is a patient that the health plan does not look for in terms of quality of care because we're not talking about cost.

Patrick Romano: Right.

Karen Pace: Right.

Patrick Romano: You may be talking about, for example, a health plan having to make selective contracting decision for a particular set of services, like whether to send their heart transplantations all to the Cleveland Clinic. And that will be...

Karen Johnson: Right.

Patrick Romano: ...from a consumer's perspective in terms of a patient with hearth failure, just wanting to know to get the best care for heart failure.

Female: But isn't the question so – was your two examples about heart failure, who the contract with and where you want to go for your care?

Karen Johnson: I still don't see the distinction. What are we looking at that would create a different metrics for one versus the other.

Female: Because it's decision making and it's decision making around the quality of care. So, are you saying...

Karen Johnson: I have trouble with this concept because I don't see any difference. And may be I'm just naïve.

Patrick Romano: Yes. Well, I think there is a difference, and you can think of it in terms of – the patient perspective is often a perspective of someone who has a certain complaint or a certain chronic illness whereas the health plans perspective is often one that is related to the need for a particular service and figuring out where to send patient for a particular service at least in a selective contracting environment. The provider's perspective is about figuring out how to allocate resources within their organization to give the best overall health outcomes to

patients that they serve. So although quality of care means different things to different actors depending where you sit in the healthcare system.

Karen Johnson: Right. But Patrick, just to take your example further, so if you're talking about congestive heart failure and quality of care and you're talking about patients choosing what hospital they want to go to, the health plan deciding which hospital they want to contract with or a provider deciding looking at how they compare to other hospitals and where they want to focus improvement effort, what would you do different in those three scenarios that, you know, are you talking about you would weight things differently, would you have different components. Or I guess I think that's the question, why would they be different.

Female: You hit it. That's exactly what I'm asking.

Patrick Romano: Yes, both of the above. Exactly, right. Because it's a different perspective to be a patient with heart failure and you don't know what the right treatment is for you. All you know is that you have heart failure and you're trying to side, well, what's the best hospital for me to go to, you know, typically when I have heart failure...

Karen Johnson: Right. But that...

Patrick Romano: ...effective than a health plan or a doctor who knows that the patient needs a particular treatment and the focus is entirely on the outcomes of that treatment. Because the decision has already been made that the patient needs that treatment you see. So it's a different – it's a less patient-centered approach because it's focused on patients who are deemed to need a particular service.

Male: There is also another argument which I think was in that paper which we hear from (inaudible) perspective which says that we might want to publish quality measures that emphasize things that are less visible to the patient. So you don't publish a quality measure about whether the parking is good, because the patient knows that the minute considered driving somewhere, but you do published a quality measure on something that the patient can't see and that could go to more subtle differentiations as well. So again from a decision-

making perspective you might – that might (expect) what you choose to include.

Karen Johnson: Patrick, this is Karen Johnson. I was just – I was thinking back to the meeting. At the beginning of the day, I really thought that you guys were distinguishing between the quality construct and the purpose and then later on during the day I wasn't so sure. So can you tell for me for sure are there two separate things? And then what you're talking about how is that different from purpose if they are two different things?

Patrick Romano: It is the same as purpose.

Karen Johnson: OK. So they are not two different things.

Patrick Romano: All right. I mean I see it as different than simply what's the quality construct is. But yes, I mean purpose I mean fundamentally in order for a measure to have a purpose, someone has to use it to inform some decision making otherwise it doesn't have a purpose. So yes, those are I think two different ways of describing the same concept.

Karen Johnson: But not the quality concept. The purpose and decision making are the same thing but not the quality construct is different?

Patrick Romano: Well, they were overlapping. I mean the purpose follows from the quality construct, but it's not necessarily, if somebody just says just like Jim's example, if somebody just says, "Well, we want to create a quality construct prevention and we want to put together these 12 different measures of preventive services." OK, that's a quality construct.

But the purpose question gets this so what. Who is going to use that? How is it going to be helpful? And depending who would use it for what purpose that would affect how we interpret and how we (inaudible) construction of the measure and the waiting and so forth.

David Shahian: The construct defines how the individual component measures relate to one another...

Karen Johnson: Right.

David Shahian: ...in some way, which is distinct from the purpose.

Karen Johnson: OK.

Patrick Romano: David, because your work with STS on the composite is illustrative here, can you explain a little bit more about how you folks thought about the purpose for your composite and how it informed the decisions that you made in the analytic methods?

David Shahian: Well, I think as I mentioned in the meeting, we (complacently) tell that mortality did not give a very comprehensive view to the patient of the totality of the care that they were receiving. And that one had to look more broadly and it's not only whether they live or die, but whether they experience the complications, whether they receive evidence-based care.

So the desire was to provide a multidimensional view of care and at the same time to increase the reliability by adding really 10 additional measures essentially to mortality so that we are better able to distinguish among providers. So that was kind of – that was the purpose behind it.

Patrick Romano: So you didn't envision this is something that would be more informative to individual consumers or caregivers in making their choices in the market?

David Shahian: And would better discriminate levels of quality, yes.

Patrick Romano: Yes, good. OK.

Female: So I'm still, you know, I don't want to be the dead horse here. But I don't know what you would have done differently if you were trying to find the quality measure that would distinguish providers for your patients. I will leave it at that. I am too dense to see a distinction.

Elizabeth Goldstein: This is Liz Goldstein. One thing, we've done a lot of testing with consumers on quality measures and so for a consumer they tend to want things rolled up more than a provider was or someone that's used to looking at quality measures. So often when we do testing with consumers and they see

lots of measures, you know, some of them are composite it's overwhelming and they don't even know what to do with it.

So if you're developing a composite for consumers, it really needs to be rolling up and then may be rolling up, you know, different experience measures with clinical measures with variant measures and (inaudible) for a lot of it were not rolling up to that, you know, higher level. But that's really what a consumer wants when you talk to them. They really don't know how to interpret, you know, a specific clinical measure even a clinical measure when it's rolled up to, you know, a few constructs rolled up together into a composite. The average consumer even your average informed consumer is overwhelmed with that. But if it is for provider they're going to (inaudible) on that detail.

Female: That's helpful.

Karen Johnson: I'm hearing somebody typing, but may be we can move on. I think, Patrick, certainly we can add some, you know, obviously we've just given you the barebones and I think, you know, the intention is the report will actually have some more context and I think the point you're making is that purpose can include this decision making context that you're talking about.

Patrick Romano: Right, right. And may be s we go into the next section, may be you could orient us a little bit onto the evaluation criteria. I wasn't quite sure about the difference between the blue colors and the red colors and would you orient...

Karen Johnson: Yes.

Patrick Romano: ...us a little bit how that table is put together and how does it get colors.

Karen Johnson: Yes, yes, yes. So the left-hand column is the updated NQF measure evaluation criteria and we wanted to have that there front and center in context because the original composite report was kind of two iterations ago in terms of NQF measure evaluation criteria. And so the red additions are potential ways we could address composites within our evaluation criteria and there are (inaudible) couple of places there. But the middle column is the composite criteria that came out of the 2009 report. And what we've done there is try to

red line or blue line, whatever the color may be, changes based on the discussion.

So what we can do is go through these. But for example in this first section, in the old report, we talked about, you know, the individual measures should be either NQF endorsed or assess who have met all the individual criteria. And as we've talked about, we're no longer recommend or you all are not or no longer recommending that and that what we're saying is that those will be addressed in the following guidance.

So this is just our attempt of, you know, taking what was in there and clearly demonstrating what we're thinking we heard in terms of, you know, either modifying or removing this. And also I should say that in that prior report, some of the things that were in the composite evaluation criteria were just reiteration of the main criteria only saying, yes this also applies to the composite. And what we prefer to do is just leave the main criteria stand and then, you know, just referenced that. But, you know, this is our first attempt and certainly, you know, I'm sure that you all will have quite a bit of suggestions here.

So the reason we have the second one crossed out is that the component measure determined to meet the important criteria then the composite would meet 1a, 1b, and 1c. And you all had a very distinct conversation about which of those apply to the component and which applied to the composite performance measure and so we just thought both of those first two things should be taken out and substitute it with what we have that follows. So the evidence criterion which is just say, you know, in our latest iteration of this we're suggesting that the evidence criterion be looked at first. So I know you're used to seeing that as 1c but this is the format that we're going to be going to in general for measure evaluation.

But basically – so the first one is the evidence criterion must be met for each for each component and of course if it's already NQF endorsed that would be given because that would have already passed that criterion. And then what you see in the left column is what the actual language of our current

evaluation criteria are about evidence. So I'll stop there and see if that helps or if there's still confusion.

Female: Currently, these are does or doesn't?

Patrick Romano: So Karen, if something appears in black in the middle column then it's carried over from the previous guidance, is that right?

Karen Johnson: No. We're saying that this is what we heard from your discussion at the in-person meeting. So the question is, is that accurate or what did we get wrong?

Patrick Romano: And what's different between what's crossed out in blue and what's crossed out red?

Karen Johnson: If it's in that middle column, it's probably just that we should have made all of the cross outs the same color. It's probably that's working on different computers. It means nothing.

Patrick Romano: OK, thank you.

Karen Johnson: It's an automatic word feature that we should have turned off and made everything out of the cross out with the same colors, so sorry about that.

So let's just go to the first one then. So what we heard is that the component, you know, whether they're just numerator components and all are none or whether the components are full performance measures that they should be evident spaced and meet our evidence criterion.

Patrick Romano: Yes, I think so.

Karen Johnson: All right, OK. So lets move on and then we can always come back.

So then for performance gap again, what we heard is that that should be met for each component and if possible for the composite measure as a whole. But then you had some discussion that if a component measure has little opportunity for improvement or little performance gap, justification for why it should be included is required.

So you talked about it may increase the reliability as a composite even though there is not a performance gap. It's very much consistent with the clinical evidence, but the idea would that they would have to provide justification.

Female: That sounds reasonable.

Karen Johnson: OK. All right.

Male: ...our discussion.

Karen Johnson: OK. And then the last sub-criteria under importance is about high priority, high impact, and the discussion that you all had with that this would apply to the composite performance measure as a whole that this you want to be doing this for a high impact, high priority area just like any of our performance measures. There's probably won't be much discussion about that but anyway that's how it would shake out.

Male: So in general, I think the theme was to reduce the extent to which composite measures review is fundamentally different measures that have to meet fundamentally different standards.

Karen Johnson: Correct.

Male: But in general, they should be viewed as measures that have to meet the same standards as any other measure. There has to be a little bit more attention to how to put together. But in terms of the overall characteristics of the measure they should be evaluated just like any other measure.

Karen Johnson: Right. OK. So moving on then for 1d, this was from the prior composite language and what we heard from your discussions at the meeting is to kind of reword this a little bit or get a little more detail is that the following must be clearly articulated for composite performance measure, the quality construct, the purpose including how the composite measure provides the distinctive or additive value and better achieve the purpose and do the components individually and you may want to talk about that a little bit more.

And then the third bullet that you talked about in regards to this was how the methods for development and the components that are used to construct the composite are consistent within representative of the stated quality construct and purpose. And we have some questions about that third bullet of shouldn't this really be what are they're going to be demonstrating in testing in scientific acceptability. But let's see what your thoughts are about what it is you want.

You know, I think we've heard loud and clear that they need to clearly articulate the quality construct and then the purpose. And, you know, how much of this other language we need here is the question for you all. And then the thing about the method, you know, our question is that we'll be getting into some of that later on is that really about this scientific acceptability when you start talking about methods. So I'll stop.

James Chase: This is Jim. One question I had is, if we have this, should we somewhere else at least define quality construct to what we mean by that, give people a little guidance about what is associated for the developers or they're submitting something, what the expectation is about, how detailed that needs to be? Or is it better to just leave it open?

Karen Johnson: So that's a good question. So some of what we've done is we put in notes to the criteria where we define things. And the other way that's some of this will be further operationalized is how we word questions on the submission form so we can, you know, do it in a footnote to the criteria and/or in how we word the questions for the submission form. But I think that's a good point that it probably needs one or both of those.

Patrick Romano: Good idea. Thank you. So if this third bullet were not right here, would it be placed later under...

Karen Johnson: Right. So we can – so if we jumped in to reliability and validity...

Patrick Romano: I mean I guess what I'm getting at is I think that the third bullet belongs there because we want people to explain how the quality construct led them to include the things that they included then they may later back that up presentation of reliability and validity.

Karen Johnson: OK.

Patrick Romano: But I think it would still be helpful to have a conceptual explanation of time.

Karen Johnson: OK.

Male: Yes. I was tripped on this one too because I was thinking about I wanted that third bullet point but I was thinking just around the components and I may be reading this different than those. If I had read this before this discussion and it said methods of development, I would think you'd want some comments about sort of how we, you know, could we get a group together to talk about this and what input they gave us. That's really not what we're looking for here...

Karen Johnson: Right.

Male: ...so I was – I was thinking it might be clear to me if this one was just about how the components are used to construct – that are used to construct the composite or consistent. And the methods of what I think we meant method of sub-development and sort of that method of evaluating it comes later in this.

Karen Johnson: Right. Is it also that the method of aggregation, you know, may be that's you want not just the list of components but also how you put them together.

Male: Yes, at least for me that would be – that would be a point of clarification. That would be helpful.

Patrick Romano: Yes. It could be right. I mean, you know, John and I we're on this panel for Leapfrog and we made, I mean we collectively made an arbitrary decision to include a bunch of process measures and a bunch of outcome measures and to give half the way to the process measures and half the way to the outcome measures. And that was pretty arbitrary choice, completely arbitrary in fact. But, you know, if we were submitting it to NQF, I think we have to explain, you know, why that approach is consistent with the purpose.

Karen Johnson: OK. So yes, we can come back to where this is placed. I think we just don't want to duplicate things then having them explained that here but also then

having to explain it in another place. But I hear you that they definitely need to explain how they, you know, how the components, how the aggregation, how the waiting, you want a descriptive explanation of those things and how that relates to the quality construct and purpose. And we need to just be more clear about that and I just say methods for development, but pull those three things out.

OK. Thank you. And that actually takes the place of actually one – the one above it, yes, sorry. The cross out at the bottom after that would be replaced. OK.

So let's move on to reliability and validity. So basically, the first one is just to really be clear what the composite specifications include and I think at the meeting nothing else was added. But if there's something else that you think the composite specifications need to include, we should add it in now.

And then the next one...

Male: May be...

Karen Johnson: Yes, go ahead.

Male: May be a little more detail. Some of these things are (inaudible) the waiting only makes sense in particular (weight) by these things standardizing, so on. May be some of the detail about the methods of combining scales into a composite should be left to the guidance rather than be put into criteria.

Karen Johnson: You mean describing the method should be in guidance versus criteria, is that what you're saying?

Male: Yes. In other words the composite specifications should indicate how the (scales) will be combined and that implies to all of those other things.

Karen Johnson: So instead of making it really explicit here, put it more in a note?

Male: Yes.

Karen Johnson: Because something – obviously, sometimes waiting won't apply or (methods of combining scales) wouldn't apply.

Male: Right.

Karen Johnson: OK.

Male: I think (inaudible) missing data and requires sample (inaudible) are distinct topic, but the way the waiting and the (inaudible) into related pieces which may or may not come up in application to see what you're doing.

Karen Johnson: OK. OK.

Male: Some just (inaudible) tested gives oral specifications of the rules of creating the combined measure.

Karen Johnson: OK. OK, so then we get not reliability testing. And I think this may go back to some of the principles and see what your suggestions are for how this should be stated. But what we had been hearing is that for composite performance measures, the reliability must be demonstrated for the composite measures for. And If the components will be disaggregated then reliability for the component measures must be demonstrated and then of course unless they are already NQF endorsed.

But we wanted to make sure that that is what you wanted to stay and so our question is, are there any circumstances where we would accept reliability of data elements, reliability of the individual component and not require or insists that they do reliability of the overall composite.

Female: I can't imagine a situation like that.

Karen Johnson: OK. I mean that's what we heard. But we just wanted to make sure we run the right track here.

Patrick Romano: Yes, I mean we're in a framework where NQF is evaluating measures for public reporting of the performance of accountable entities. And so it has to be shown that there is, as Sherrie said, some prints that there is a recognizable

signal of these accountable entities. So I think that's generally what we mean by reliability in this context...

Karen Johnson: Right.

Patrick Romano: ...and it doesn't make sense to have the measure be NQF endorsed to public reporting without that.

Karen Johnson: Right. And just to clarify and this is also something that comes up in another project. But, you know, in our current, you know, evaluation criteria as the results of the measure testing task force we've allowed measure developers to test reliability for either the data elements or the performance measure. And, you know, I think, you know, that whole framework may need to be looked at as well in kind of heightened interest and accountability measures. But I think this is reasonable and we just wanted to make sure that this is indeed that we captured your recommendations so.

OK. And then on – go ahead.

Patrick Romano: And this gets to I think it's the topic that we did discuss a little bit. So if the components will be disaggregated then reliability for the component measures must be demonstrated. Now, the question is this is aggregated for whom?

Karen Johnson: All right. So this would be disaggregated for public reporting or for accountability, right? Or what...

Patrick Romano: Yes. Is it OK to clarify that? Because in some cases, we may want to encourage desegregation...

Karen Johnson: Right.

Patrick Romano: ...for provider...

Karen Johnson: You're right, yes.

Patrick Romano: ...some data.

Karen Johnson: Right. I think that's reasonable. Yes, so we can do that, yes.

Patrick Romano: So this is the situation where, you know, I think we're basically saying that if you're disaggregating the components for public reporting...

Karen Johnson: Yes.

Patrick Romano: ...then the components basically have to meet the reliability, validity standard that they would meet as separate measures.

Karen Johnson: Right.

Patrick Romano: But there is separate space for components that may not be separately disaggregated for public reporting.

Karen Johnson: Right.

Patrick Romano: OK.

Karen Johnson: OK. All right. And then on validity testing to B2 for composite performance measures, validity must be demonstrated for the composite measure again. And if the components will be disaggregated then validity for the component measures must be demonstrated unless they are – again, we can work on the wording there it's a little clunky. But the idea is the same thing that you just said for reliability that if they're going to be disaggregated for public reporting or accountability then they need to meet the criteria to stand alone, right?

Patrick Romano: Well, this may need a little work and I want to hear what others think. But because in many cases the validity measurement does come from the individual component intrinsically, right? So if we put together five process measures then for each of those process measures we might have information that that is a collected validly. So then we want to put together those five measures into a single composite. If there's some further validity testing that we need to do on the composite or is it sufficient to say that each of the five components has demonstrated validity.

Karen Johnson: So I think that's a good question and I think, you know, it seems like perhaps it went back and forth that I mean, because part of it is, you know, how you end up aggregating these things. You know, so yes may be they were valid as

individual performance measures but then for some reason the way they all got put together into a composite, is it so a valid indicator of quality or, you know, and how do you know that.

So it's a good question I mean and so is this just the opposite that if the individual component would determine to have validity that you don't have to worry about that for the composite or do you because you're really doing a lot of manipulation with that data to get it into a composite.

Patrick Romano: I think the illustration you just gave is exactly right and I think I can't think of one right at the top of my head but I think it would seem to me perfectly reasonable to think that there might be situations where you might take half a dozen valid individual indicators and combined them in such a way that the resulting composite was no longer valid for the purpose it was designed. So I think that's a valid point.

Male: (Inaudible) you know, a few weeks ago was the various kinds of validity. So there are quite a few ways that something could be shown to be valid and (inaudible)...

Karen Johnson: Right.

Male: ...admission of that but there should be at least one of those should be there.

Karen Johnson: For the composite.

Male: For the composite, right.

Female: It's entirely possible that the weighting for example and equal weighting over several measures some of which have much lower priority than the others would not be a valid reflection of quality. So I think we have to have the composite reflect quality and demonstrate that in terms of validity.

Male: OK. Well, I like all that. But how do we give specific guidance to measure developers about what analysis they should do or what information they should present to support the validity of the composite as a whole?

- Male: I think that gets back to having some (textbooks) summarizes the various versions – various approaches that’s showing validity which are not necessarily in convergence with each other that we discussed in the meeting.
- Female: OK. We can go back and pull some of those out.
- Male: Right. What – I mean one example, for example, is when we look at individual components, we often rely on criteria and validity where we, you know, invoke some kind of a gold standard and we ask whether the data collected here are accurate according to that gold standard. But when we’re looking at the composite as a whole, we may be more interested in (inaudible) validity. We may be more interested in convergent validity with other concepts...
- Female: Right.
- Male: ... that represents relevant quality.
- Female: Right. Good point.
- Male: And Liz previous mentioned, you know, that a lot of it comes down to correlation with outcomes. So, if we’re putting together a bunch of process measures, the acid test to validity is whether this composite of process measure is correlated with patient outcomes.
- Female: Right. And what was then – that brings us the question of what would you do if it was a composite of outcome measures for validity?
- Female: Well, you can do known groups kind of thing or maybe the opposite direction of correlation with known processes.
- Male: Or you might have a set of outcomes that (commensurate) with each other.
- Female: Uh-huh.
- Male: You know, like survivals for very exclusive patients that are – that have this similar but not exactly the same condition but different treatment.

Dale Bratzler: So, I – this is Dale. I'm a bit concerned about tying in any way the validity evaluation to outcomes. I guess some people could do that, but I think we have to be cautious. For many of these process measures and composite measures when you roll them up, they still represent a fraction of care and to expect people to be able to demonstrate that performance on a composite actually impacts outcomes or other things. We just have to be careful there.

Male: I think that's just one of a number of options.

Dale Bratzler: Yes.

Male: (Inaudible) set of composite measures that are grouped together just because they correlated with each other. There's very little information in putting them together and you wouldn't – as long as they were each individually valid, you wouldn't need to move that. The composite was a better predictor of outcomes.

Female: I also think there has to be some outcomes that are impacted by these things. So, why are we collecting them?

Male: But that evidence might come from prior studies which are cited in 1A.

Female: Right.

Male: Rather than evidence of validity that's presented here.

Male: Right. The individual measures might be related to outcomes based on, you know, clinical studies or whatever. But the way that you put them together into a composite might not have the specific outcome for which you have data to prove. That's the right way to composite them. We might have other grounds to composite them that way.

Male: Yes. I completely agree with that. I mean you can take all sorts of composites that are already out there with things like diabetes measures that are grouped together. You know, to demonstrate improved outcome, you may have to follow the patient population for years.

- Female: No. I don't think – I don't think it was to demonstrate improved outcomes. The idea would be that if the theoretical framework is if you have a group of process measures then you're doing, you should have a correlation with the outcome measures. They should be moving in the same direction not – I don't...
- Female: (Distinctly).
- Female: Yes.
- Female: And there have to be some outcomes. For example, hospitalizations or...
- Male: So, I agree with that. And I think that's actually easier at the level of the individual measures where you may have, you know, randomized trials or other things to actually – to make that association but with composites, I think, Patrick or somebody mentioned, you very likely aren't going to find those types of studies that have clearly linked the composite to the (inaudible).
- Female: Right. We're not asking for the clinical studies. This is strictly about the performance of the measure in a construct validity standpoint.
- Female: It should be able to demonstrate correlation with outcomes in data.
- Male: I think that's too high of standard.
- Male: Yes. I don't think you're going to be for many composites. Even one says are out there and well accepted, you're going to be able to clearly show correlation with outcomes.
- Female: So, you're saying there are – there are composite measures out there that are being used, people are putting in resources to collect these composite measures and put them together, but they don't have a correlation with outcome?
- Male: I think many of them may well but demonstrating it is something that may take years.

Male: The practical question is let's say you put together a bunch of process measures, each of which individually comes out of randomized trials showing some benefit in terms of outcomes, then when you create a composite from these measures, do you further have to show that the composite is correlated with patient outcomes and presumably that correlation may be stronger than it would be for the individual component measures by themselves.

Female: That's what I would hope.

Male: You would hope?

Female: Yes. I mean I really – I mean – I'm sorry. I just think that there have been a proliferation of measures and they have to have a purpose. And if they're not collected for the purpose of demonstrating that outcomes are better, what is that purpose? Or for the purpose of...

Male: I'm just saying that those evaluations are very, very difficult and often times take evaluation of populations over long periods of time and recognizing that any composite is made up of individual performance measures each of which have some of them substantial population where patients were excluded. It's just difficult to do that linkage. So, I'm not opposed to having construct validity. I think that's very important. And I think it's – I think it's easier if the individual process or individual measure – individual component of a composite to demonstrate based on published evidence of others some linkage to the outcome of interest. I think when we start rolling them up it's difficult to say that you're going to find the evidence that you need if we end up throwing in something about a linkage between performance on a composite and outcome.

Female: Right.

Male: So, I don't think you're going find many studies for many measures.

Female: So, I need to make a clarification here. The clinical evidence between a process and outcome is the clinical evidence that we ask for under importance. For any kind of measure, whether it's a composite or a single process performance measure, what ideally we want under validity is construct

validity on that measure as constructed, and what is the construct validity, you know.

Again, you could do the process measure with outcome measure. You could do process measure with other process measures that you think should be moving in the same direction. There's a variety of things based on the conceptual framework, but the distinction we make is that the evidence under importance is about the clinical evidence and studies.

What we're looking for under scientific acceptability is the measure as constructed. So, whether you're talking about an individual performance measure or composite, how you put that measure together can affect the validity of what you're measuring.

Male: Well, this is undeniably true. Is it possible to set here a kind of – two different bars? And I know that some of the scoring that (Jim) – that the steering committee had to use gets a little complicated, but ideally we would see evidence for the composite as a whole being predictive of or correlated with outcomes that matter to patients and other stakeholders. But in the absence of such evidence, then committees would look at the evidence for all of the components of the composite individually and typically it would be about then the criterion validity of those components.

And in some cases, that may be sufficient, but certainly a steering committee could say, "Well, you know, we'll let you get away with this for a few years, but when it comes back for maintenance, we'd like to see that higher standard of evidence of validity." Perhaps, there's (inaudible) that kind of approach.

Female: Perfect. Yes.

Female: Right. You know, going back to what our criteria are for any measure is we still have this out of face validity. So, even the individual performance measures – you know, I don't know what my guess would be and what percentage of them have actual empirical validity analysis versus relying on face validity because of the clinical evidence that exist. But if that's what our question is, what are you recommending would be acceptable if you think that's too high a bar to say that empirical validity testing is too high a bar?

So, when you say the individual components, are you willing to accept face validity of the individual components?

Male: I thought we already had the criteria for the individual components as being that they're essentially those of an endorsed measure.

Female: Well, that was only if it was going to be disaggregated for public reporting is what our understanding was and actually, you know, right now, our – that's what I'm saying. Right now, our basic criteria would allow face validity that individual performance measure – and I'm just saying, you know, we're coming long ways from saying you want empirical validity testing of the composite to accepting face validity of the individual components, and you know, we can do that. I'm just – I just want to be clear on what we should capture.

Male: Well, can you clarify how (2B2) is interpreted in the context of individual measures?

Female: Right. So, well, if you go on to – actually, I think we're – I'm missing something or maybe it's in – let me look in the note 11 because we do talk about – yes. If you look at page 17 – I just didn't want to make that particular table go on and on, but the note 11 that goes with validity testing, at the end of that note – do you want to highlight it? Can you highlight this statement about face validity?

The face validity of the measure score is the quality indicator maybe adequate if accomplished through a systematic and transparent process by identified experts and explicitly addresses whether performance scores resulting from the measure specified can be used to distinguish good from poor quality. So, you know, again, the criteria is about validity testing but we do still allow for face validity at this point in time, you know, that may certainly be something that changes eventually in the criteria, but...

Male: Well, this is probably not the time and place to revisit that paragraph.

Female: You're right.

- Male: I mean it does make it clear that face validity is sort of marginal...
- Female: Right.
- Male: ... maybe adequate. So, that certainly leaves room for steering committees to decide if it is not adequate.
- Female: Right.
- Male: And I don't think that this is the time and place to change that rule.
- Female: Right. So, the question is whether you wanted to set a higher bar for the composite measure or just go with the basic criteria as we have them. So, what I'm hearing is that at least some of you want to – obviously, the ideal is validity testing of the composite measure. The ultimate would be validity of the individual component, and in that case, it would default to our current criteria which would allow for face validity in some circumstances of this individual component.
- Male: That's right.
- Male: I'm comfortable with that.
- Male: Yes. I mean it's – I mean recognizing Liz's point, which I agree with, you know, we just need to show a fast forward, you know, that – that this is the ideal standard, you know, for what – what we would like to see measure developers submit. But in practice, it – it may not always – it may often be impossible for them to submit that but we would still like to see them working towards that – that goal because ultimately it's what matters.
- Male: I sort of miss the last statement that – that (Karen) made. Is it – are we saying that it is now possible to have some sort of testing of the individual components for validity including a solidity but not test the overall composite.
- Female: That's what I was hearing, that people wanted that as an alternative to validity testing of the overall composite.

Male: But what about getting back to how this discussion originally started, that it would be perfectly possible to have some (inaudible) waiting scheme of the individual components of a composite that made the resulting composite measure itself, no longer a valid indicator of quality.

Male: Well, yes that could happen but how would we operationalize a specific instruction to measure developers, so let's – let's think about a – a composite of outcome measures. Each of the outcome measure will assume has been separately validated, so how would you validate a composite of this outcomes? What would be your method that you would look to measure developers to use?

Male: So – so Patrick, this is John, I agree with David's last point and you know in the comments that preceded that – that thing – that you know, kind that the primary goal should be to validate Dave's summary measures that the composite process ends up with but, you know, but for the large majority of measures, I suspect that (inaudible) isn't feasible for one reason or another that two things should need to be present: 1) The components of the composite should be evaluated basically by the same set of criteria that – that NQF already has in place for individual measures whether – whether they're empirical of – of test of validity or whether it's (inaudible).

But as David was indicating, there needs to be some consideration for – for the rational combination of individual measures whether it's, you know – you know, whether it's empirically based waiting or whether it's clinical judgment based waiting.

Female: So – so maybe that will – maybe we'll jump ahead because I know we're running out of time and I think what we wanted to discuss with you is some of the additional criteria for composite measures, so maybe people would be more comfortable with where we at now with the alternative if we also requiring, you know, some analysis to support the waiting or the components included, so that people have more information on which to either agree with the kind of, you know, moving forward without empirical validity testing of the whole composite or – or not but if we can move to page 9 and let me just explain what we have here.

Actually, 2I, 2L's were from the prior composite report and one of our question is, do we need each of these individual element as separate criteria for composite or can we do something that was more encompassing like 2D. I'm not saying that's exactly right but – but in light of our discussion that we just had, you may want this as separate criteria, so maybe we can go to these.

So 2I would be about component measure analysis demonstrates that the included component measures fits the quality construct. And then along with that, the next one is for component measures, either empirical or conceptual justification must be provided to demonstrate that – oh, that was about the parsimonies, one of your recommendations the other day I think related to parsimony, so that may or may not belong there.

Two K was the scoring/aggregation and waiting rules are consistent with the quality construct with the – and then here was about the preference for simplicity and easy presentation. And then 2L was about analysis of missing data. So again, you know, we can talk about wordsmithing but the idea is – for the alternative would be to have something that was more encompassing like 2D for composite performance measures analysis support that the specified scoring and waiting and included component measures are consistent with and representative of the stated quality construct and purpose and then with adequate attention to parsimony and simplicity.

So, I guess maybe one question is whether you want individual criteria that relates to the component in the scoring and waiting and totally open for discussion on which way you want us to go with this.

Male: Just to make sure I'm understanding the question, were you saying either – either 2D is enough or do we want to have 2I, J, and K, they're I, J, K, and L?

Female: Right, right and – and ...

Male: I actually thought ...

Female: Go ahead.

- Male: That it was helpful to have the – the more – to break it out into each – each one of them, so I – I was as I read through this, I thought it was helpful to have the 2 I's, 2 L's.
- Female: OK.
- Male: And – and maybe that means we don't need 2D but I ...
- Female: Right, right. I think it's an either or. I mean we could, you know, we can get more detail in 2D but you know it – it's really just a question for all of you what you think is the best way to go.
- Male: This seem to be reasonable to me when I read them based on our conversation of – of a way to describe what would be looking for and I think the additional specificity maybe helpful.
- Female: I can buy that.
- Male: We seem to have (inaudible) in 2I's but as well we seem to have lost the parsimonies, simplicity.
- Female: If you look under – well I don't know you should probably still be 2J the – the – for component measure, either empirical or conceptual justification must be provided to demonstrate that adequate attention has been paid to parsimony. I don't know if that should be combined with something else but that's where we have it right now.
- Male: OK, I – I guess I'm (inaudible) if they are left out or – or is that becoming part of 2I?
- Female: Right, it can either be part of 2I or we can leave it as 2J.
- Female: Lisa just highlighted the part that we're talking about now.
- Male: But I – I don't think that – it should be may be something about component measurement. It's about the (inaudible) combination of the parsimony and simplicity component.

Female: OK.

Male: ... (Inaudible) of component and – and the way that they're put together.

Female: Right, so 2K, the scoring aggregation and waiting rules are consistent with the quality construct with the preference for simplicity. So, I – and maybe we're wrong about this. We saw parsimony related to the number of components and simplicity related to the whole scoring rules in aggregation.

Female: That sounds right to me.

Male: Yes, I'm getting it. I'm just confused by it's (inaudible).

Female: Right, sorry.

Male: So – I – you raised a marginal comment, what would be the (inaudible)?

Female: Yes.

Male: And – and I think, it looks to me like the original version of this was focus on the psychometric approach of – of correlational analysis and when you crossed that out it's not – it's not clear what analysis mean. It could be analysis might just be an enumeration of – of the common – of the effect and a common outcome for example.

Female: Right.

Male: So, I don't know if the analysis is exactly the word there.

Female: OK.

Male: It – it maybe it's justification which could be an – a quantitative analysis or it could be more based validity based on looking at the known properties of each individual component. (Inaudible) of eight things that all affects mortality but there's no quantitative analysis of exactly how to put them together.

Female: Right. And – and we have this criterion about parsimony, what again, you know, I – I think this is, you know, gets to your recommendations about examples on – those kinds of things but, you know it will kind of beg the

question is what would be the empirical or conceptual justification to show that you've addressed parsimony. So, empirical could be, you know, if it doesn't add anything to your model but I just wanted to see if we can provide more guidance here, what we would need to do for example of what could go in notes or in the text.

Female: So if two measures are highly correlated and almost identical certainly.

Female: Right.

Female: They would not qualify.

Female: Right. OK and I know this a lot to throw at you, so right now what I'm hearing is that in favor of having the separate sub-criteria that are specific about composites and obviously we can work on the wording a little bit. What we'll do is kind of clean this up send you another version that you could redline or make any wording suggestions to us.

Male: I mean, I think what's important here is it just needs to be reframed a little bit so they sound a little less like a hardened, fast rules and a little bit more like you know, here is what the steering committees will be looking for. So, if a component measure is not at all correlated with the other measures, then there needs to be some clinical justifications, some space validity justification if you will for why it belongs.

Female: Mm-hmm.

Male: If it is extremely highly correlated with the rest of the composites, then there needs to be some just justification for why it can't be taken out or why...

Female: Yes, right, OK.

Male: So, I think that's what we're looking for. It's really what the steering committees will be looking for in terms of the evidence presented.

Female: OK. And I think we'll kind of shortcut them – the feasibility and usability, and that this basically is the main NQF criteria applied to composite performance measures. Again, we'll give you a kind of a non-redline version

so you can more quickly look at it to see if you have any issues and we'll get that out to you pretty quickly.

But I think that maybe what we can do in this interest of time is go down to additional recommendations and maybe additional questions if we have time. But, the other thing that we heard was about your recommendation about the review process and steering committees should have at least one member knowledgeable about composite measures.

And certainly, your recommendation to provide examples and types of analysis for different types of composite measures is great. We'll have to figure out a way to do that. I know that at one point, you suggested commissioning on paper. We don't currently have funding for that but interested in any of your suggestion on how we might start working on that. And then – go ahead.

Male: Just a suggestion. I didn't necessarily mean (inaudible) the staff could probably do a pretty good job and get some inputs and (inaudible) to review it.

Female: Right. One of the things we thought – I mean we were talking a little bit about that is trying to out at least together one example, and then we did you know, obviously continue to work on that but we would've definitely need to come back to all of you for some review and suggestions on that.

And then the one that we came up with today was about – in terms of process is giving developers an opportunity to rework a measure during the project timeline, you know, specifically related to this idea if a particular component wasn't you know, looked upon in favor not to lose the whole composite measure if the developer was willing.

So, are there any additional recommendations that you know, you all thought of during your meetings that we didn't capture here that we should be sure to do.

Male: There's one issue that we, I think are a little stuck on at the end of the meeting which was how you're going to formulate the criteria of what a composite is so that you know which submissions are channeled into this process.

Female: Yes, thank you. We skipped right over that whole definition thing.

Male: I thought about a little bit on the way home, and I thought maybe better than trying to have a you know, big super-duper (inaudible) everything definition of the composite measures, it might be better way with systems to have a series of questions that sends people off to provide the appropriate information.

You know, (inaudible) characters it's like, you know, if this combines NQF approved measures, then you go to one place to submit those in and reference them if it brings them some more new measures, you go somewhere else to get information of the new measures and then if there's a combination of distinct measures that are (inaudible) the way (inaudible) talk about why you're combining the way you did or something like that.

(Inaudible) trying to have a big, broad definition of composite measures that applies for everything else. I don't know how that works with your systems.

Female: Right. So, going back to the definition, I wanted to revisit whether we need to change this. The performance measure that's in brackets was actually an addition from the original and just thought it would be good to highlight. We're talking about performance measures versus a composite like the – or you know, what some people referred a composite at the kind of individual application level.

But we didn't hear a lot of discussion about changing the definitions but we wanted to revisit that with you as well. And then, of course as you just brought up, how do we actually operationalize that when measure developers come forward of self-identifying they have a composite and especially in relation to some of these additional criteria we were just talking about don't necessarily fit the all or none composite.

Obviously, they still need to justify the inclusion of the component of why, you know, the evidence for each of the components, and why the rationale for having that is an all or none but I'm not sure that the item, you know, that component analysis makes sense in that context.

So, it gets a little tricky if not only it's a composite but then which questions need to come up. So, I don't know that we can solve that today but if as for the first question is whether you want to alter the definition at all.

Male: One thing we did struggle with this but I think where we agreed was that it was very hard to draw a bright line between our composite measures and other types of measures.

Female: Yes.

Male: And so, we sauntered a little bit in our efforts to do so but if we stick with the concept that NQF is endorsing performance measures, and therefore a composite performance measure has to be a composite of individual performance measures.

That seems to be a reasonable foundation for some branching points as Allan suggests that would kick in as measure developers are responding. So, instead of being a completely separate submission process, it's all part of the same process but there are some branching points as Allan has nicely illustrated.

Female: OK.

(John): So, this is John and I don't mean to backtrack but is there an important distinction that's being made between performance measures and quality measures? Because all of the component measures I think are defined in terms of quality, right rather than performance?

Female: Right. I think you know, we used – its quality performance measures that's said to be fully – but you know...

(John): (Inaudible)

Female: Pardon me?

(John): So, they're synonymous? Is what you're saying.

Female: Yes, right.

(John): So, OK, never mind.

Female: We are talking about performance measures on quality. The reason some of our language has gotten more generic with performance measures is the (inaudible) strictly quality is moving into the space of efficiency and resource use. So, anyway, we are meaning the same thing.

Male: Yes, I mean what we're just trying to stay away from is the situation with some of the caps domains or some of the functional status domains where the methodologists would tell us that they simply need to have three or four items in order to measure a single concept and each of those items you could call a measure if you want but it's not really a performance measure because no one's arguing that it could be used to evaluate the performance of doctors or plans or hospitals or anything else, one by one.

Female: Right. And you know, to complicate things, you could take all of those different domains from caps and create a composite performance measure as well. But it gets, you know, we won't go there again. But you know, it's just gets confusing with...

Male: (Sharon) used the term high order composite.

Female: OK.

Male: And that's you know, maybe it's a useful term. I don't know.

Female: Right. OK. All right. So, we'll go back to the – are there any additional – oh, yes. I'm sorry. We need to pause and see if there's any public or member comment.

Yes, (Madelie), could you open the lines for us for our members in public to make comments on the discussion.

Operator: At this time if you would like to ask a question or make a comment, please press star then the number one on your telephone keypad. We'll pause for just a moment to compile Q&A roster.

Again, that's star one to ask a question or make a comment.

At this time, you have no questions or comments.

Female: OK.

Male: So, with respect to the second and third bullet points under additional questions on the (inaudible). So, perhaps people may want to provide some feedback in writing.

Female: Yes.

Male: Certainly, my first impression is that – to the second bullet point is yes, this is a recognized problem with all or none rating is that it may not achieve the goal of increasing reliability depending on that particular distribution and co-variant matrix of the measures.

So, this is something that perhaps should be a topic for empirical analysis and discussion. But I don't know what more to say than that.

Female: Right, OK. The reason you know, it comes up sometimes because you know, as you all have talked about one of the reasons for the composites and having multiple measures is that you increase reliability because you have more data points.

And we've heard that rationale applies to all or none because there are more components in enumerator but when you reduce that down to basically one day to point either yes, they all were met or no, it was just, you know, a very specific question we know but we're just interested in your input and certainly can do that offline.

Male: Right. Well, (Reeve's) analysis and others certainly suggest that often you lose reliability.

Female: OK. OK.

Male: One item with a very lower performance and you've thrown away all the other information.

Female: Right, OK. And then I think we know this and Dave (inaudible) can certainly speak to it and probably Patrick but you know, often times we see the discussion about the analysis, the factor analysis and internal consistency and reliability when you're talking about items of this scale, and just wanted to make sure that you know, when we talk about this in the criteria, that (Reeve's) analysis also work when you're talking about scores on a performance measure and the units being you know, organizations versus the individual people answering a survey, for example.

And our understanding is yes, you can apply those but just wanted to confirm that so if we put those in this example.

Male: I think you have to mention that the analysis of that type conducted at a higher level unit level has to take into account reliability of the measures and this is an area I've spend a lot of work in this, it's not an obvious fix. It's (inaudible) reliability measures and you're trying to assess how related they are.

Female: OK.

Male: Could you pause – Allan, I think it might be helpful if you write a few sentences along those lines?

(Allan): Sure.

Female: Yes, that would be great.

Male: And similarly for others I think everybody should feel comfortable, I probably will too to suggest some sentences for a background or further elaboration or explanation.

Female: Absolutely. So, and you know, we appreciate you being in this with us. We're on an extremely tight timeframe so we wanted to get kind of this you know, basic recommendations that we definitely need to put some narrative around this and we're you know, -- so any suggestions about other contents that should go in the report and references or in the appendix, you know,

would be very appreciated of any suggestions on wording and additional content.

So what we will do is you know, pretty quickly get through these principles and criteria in a cleaned up version without any redline so that that's what you can work up off. And let you know when we need to hear back from you. And certainly, you know, if you think of anything after the call or at any point, you know, shoot us an e-mail and we would really appreciate your help.

Female: Once again, you've done a terrific job corralling all these texts.

Female: Well, we appreciate it. It was useful to have all of your thinking to work with. So, we thank all of you. And (inaudible) if you want to make any other comments or suggestions, we're ready to do that.

Male: I don't have anything further. Thank you very much for your staff leadership in this, and again, I just encourage my colleagues on the committee to put some thoughts down on paper or on e-mail, and share them with the group.

Female: All right, well, thank you all. Lisa, do we need any other reminders.

(Lisa): Just some housekeeping items. If you received – if you haven't received information from our meeting's department with regard to reimbursement for the in-person meeting, I think it was – was that last week? About two weeks ago, please let me know as soon as possible. We want to make sure that we get those into our meetings department.

And we will also send you information about the two upcoming CSAC meetings. We would of course, Patrick and Liz will be representing the technical expert panel but we would love for you to also join in and listen to the discussion.

So, the first one in December, the second Monday in December. I think it's the 10th of December and the second one is January eight. So, we'll talk...go ahead.

Male: We have one last conference call for this group. Do we?

Female: Yes, we do and that will be to adjudicate the comments after the comment period. So, tentatively, we are hoping to go out for comment with this draft support you know, as (Karen) mentioned earlier, she will send you the clean copy of the evaluation criteria and guidance as well as the draft support. This is where we'll go out for comments.

And we're hoping to go out for comment on November 28. There'll be a 30-day comment period. We'll let you know of any changes in that schedule but we have the comments call on the 3rd of January. It's scheduled right now for the 3rd of January. So, we will send you details as soon as possible for that.

Female: What date did you say in December that we'll have that conference call?

Female: That's for our consensus standards approval committee and we'll be presenting the draft report without member and public comments. As I mentioned during the in person meeting, we want to get in front of the CSAC early because we're making some changes to our criteria and guidance as it relates to composite measures.

And so, we want them to have enough time to process this information for them to make an informed decision. And so, we're going to them. We went with an overview just a couple of weeks ago...yes?

Female: (Inaudible) on that call and I'm looking for it on the calendar and I don't see it.

Female: That's why I was going to ask. I don't know if (inaudible) has set that up with Liz and Patrick.

Female: Not yet but it was part of our (inaudible) so, yes.

Female: So, we'll get back with you about that and if there's a conflict, we'll you know, we can deal with that.

Female: OK.

Female: We'll get back with you on that.

Female: Thanks.

Male: OK. And so it's a surprise.

Female: Sorry about that.

Male: OK, thanks.

Female: OK.

Male: Bye.

Female: Thank you everyone.

Operator: Ladies and gentlemen, this concludes today's conference call. You may now disconnect.

END