NATIONAL QUALITY FORUM

Moderator: Centered Care Person & Family
June 26, 2014
2:00 p.m. ET

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

Lauralei Dorian: Great. Thank you. Good afternoon everyone. This is Lauralei Dorian and the NQF team here. Thank you very much for dialing in today. Before we get started, I just wanted to remind you that this call is open to the public and it will be recorded and the transcripts will be available following the call. And also, just to remind you to please keep your phones on mute unless you wanted to speak.

So we wanted to thank you, the committee members on the call for submitting surveys if you did so for reviewing these measures. We know you didn't have a lot of time to do so and they're quite complicated and we'll talk about some of these complications this afternoon. But we do appreciate the time that you put into the call. We have developers on the call today who will be available to respond to any questions that you may have. And before we get started, I might just go through so see just to make sure we know we have on the call. I see that we have (Kimly).

(Kimly): Yes. (Kim) is fine.

Lauralei Dorian: Kim is fine. OK, great. Brian?

Brian Lindberg: Yes, I'm here.

Lauralei Dorian: Great. Hi, Brian. (Sherri)?

(Sherri): Yes, I'm going to be on for half after and then I have to jump off for half hour and then I'll come back.

Lauralei Dorian: OK, great. Thank you. Jim? OK, Deb Saliba? And Carin van Zyl.

Carin van Zyl: That's me.

Lauralei Dorian: And Len Parisi.

Len Parisi: I'm here.

Lauralei Dorian: Great. Welcome. And then can I just ask any measure developers on the phone to identify themselves please maybe for starting with the Veterans Affairs, the PROMISE Center group.

Female: (Inaudible).

Lauralei Dorian: Hi, (Inaudible)

(Joan Philbert): (Joan Philbert).

Female: Hi, Joan.

(Joan Philbert): Hi.

Mary Ersek: Mary Ersek.

Lauralei Dorian: Sorry what was – who was that?

Mary Ersek: Mary Ersek, the Director of the PROMISE Center.

Lauralei Dorian: Hi. Mary. And then from Brown.

Joan Teno: Yes, hi. Dr. Joan Teno.

Lauralei Dorian: Great. Hi Dr. Teno. And do we have anybody from the National Hospice and Palliative Care Organization?

Carol Spence: Yes. This is Carol Spence.

Lauralei Dorian: Hi, Carol.

Matthew Haskins: And this is Matthew Haskins.

Lauralei Dorian: Hi, Matthew. So I think we just did want to remind you before we got started of the difference between performance measures and surveys. All of these measures are based on surveys but just to remind you that NQF does not endorse surveys themselves. We endorse performance measures so that means that we're endorsing measures that are measuring something at a different level not the individual patient level but a level like a hospital or something that's an accountable organization. So as we go through some of these measures, we're going to want to make sure that testing is done not only at the data element level which is the survey level itself but also at the performance measurement level.

We can get into some of the details as we go through the measures. And before we get into the specific measures, I just wanted to pause to see if anybody had any reflections on how these process went for them or any sort of overarching questions that aren't about specific measures but – that you didn't quite understand the question or survey themselves in general.

OK, well with that, I think we're going to start with Measure 1623 which is the Bereaved Family Survey which is from the Department of Veterans Affairs Hospice and Palliative Care section – sorry. (Karen)?

(Karen Pace): Let's just talk a little bit about what process we're going to use today.

Lauralei Dorian: Sure.

(Karen Pace): So, what we're going to do today, we didn't ask the committee members to present a measure. Lauralei is going to kind of walk you through the evaluation criteria for each measure and we'll ask the committee members to identify any questions or observations they have with – about the measures in relations to those criterions or things that were not clear so that we can – if we need to, we can ask the developers to clarify some things and then also the

staff will add some input as we've been reviewing these measures along with you so we can also bring up any questions or concerns as we go along.

Lauralei Dorian: Are there any questions about that for today? And this preliminary discussion is really going to help and get you familiar with this process and that measure evaluation criteria that it would be a good starting point for our conversation during our in person meeting so that, we'll actually assign everybody as a lead discussion or a secondary discussion for each measure and you can begin by summarizing some of the concerns or not that were raise about these measures during our in-person meeting just to give you a bit of how's that preview.

So, 1623 is the Bereaved Family Survey. As I said, the measure assesses family's perceptions of the quality of care that veterans receive from the VA in their last month of (inaudible). It consists of 19 items. It has – we could ask to mute your line if you're not speaking.

And has one – and correct me if the developer is around the line and wanted to correct me with anything but it has one ultimate measure at the end. So it has a global item question and an optimal response on the denominator as I said is all inpatient (depths) for which the survey was completed.

Also, we're going to, (Nadine) if you could just go to the first. So this first measure I think we'll go through each of the survey questions just to make sure you understand what their asking. But for the following surveys or for the following measures, we won't go through each survey question individually but we'll ask you to sort of raise questions in the order with the measure evaluation criteria.

So, Karen, did you think I should go through the survey responses or just ask an open-ended questions with …

(Karen Pace): No. We'll …

Lauralei Dorian: We'll start with – so 1a focuses first on evidence to support the measure of focus.

So, if you look at that first bullet, it's measuring a health outcome or PRO, Patient Reported Outcome is the relationship between the measured outcome PRO and at least one health care action which would include a structure of process intervention or service identifies and supported by this dated rationale.

(Karen Pace): So let's bring up what they put in their evidence form.

Lauralei Dorian: They are going to bring up the measures submission.

(Karen Pace): So, for those of you who are workgroup reviewing this measure, this relates to what they put in the evidence attachment meaning you need to go to the evidence attachment. And so, the question is whether they provided any information about healthcare interventions, structures, services that affect or influence this particular patient reported outcome about their experience with a positive care.

Len Parisi: Can you repeat that again.

Lauralei Dorian: Right. So the question is what are requirements for evidenced related to outcomes and PROs following this context is, is there a relationship between the measured outcome in this case experience with hospice care and interventions and that the healthcare entity can provide that will affect that.

So, we ask that the developers provide a diagram or description of what healthcare services or interventions affects the outcome and give a rationale for that.

Carin van Zyl: This is Carin van Zyl. I'm unclear whether the evidence that's needed to the submitted by the developer includes what they think will be altered by the information gathered from the performance measure or what has already been improve or change as a results of either pilot data or the measure that's been in effect for sometime.

(Karen Pace): This actually has not have to be related to the specific measure but this is an overall measure of experiments with hospice care and they talk about the different domains that are included. And it's really about what are the things

that hospice organizations can do, our palliative care organizations can do to really influence what the patients experiencing.

I mean, I think it's a pretty simple bar for outcome measures in general but essentially the PRO measures. So could you get to the information they actually provided and maybe – so, we had gone through on some of the tutorials, what kind of information we're looking for here which is really basically to say, what are the – you know, is there at least one healthcare structure process or intervention or service that will effect the results on this outcome for patients? So, basically, what we are looking for is you know, an example here would be you know, prompts and efficient. Prompts and effective pain control will affect the patients experience the family's perception of the patients experience with (hospice) care. Or you know, in general we really put symptoms or, whatever kinds of things that are actually being measured within this so they and that are being compiled into the performance measure.

So, you know, what seems to be put here is really just a general rationale for why you want to look at persons experience with you know, some of these experienced items or how you would want to administer the survey. But it's not exactly what we we're looking for but I'll stop there and see if the committee has any thoughts or questions that we want to post or that you're thinking about.

Len Parisi: So this is Len Parisi.

(Karen Pace): Yes.

Len Parisi: So, in looking at this, I concluded that if evidence-based practices related to end of life care, palliative care, hospice care will implemented, it would have a positive effect on the response to these questions in the survey. It was very helpful Lauralei when you explain that we're not looking at the survey, we're looking at a measure that maybe derived from that and then the other two measures that we reviewed, it was clear that there was a composite scores so that's clear for me and very helpful because I had a related question about the

upcoming CMS Hospice CAHPS Survey that was coming down the pipe. So, that clarified it for me.

But I – the way, the real ends that I look at this, I concluded that if a patient was receiving appropriate interventions during their end of life that there would a positive effect on the responses to the survey and a result on positive outcome. That's how I interpreted it. But I did have a question and maybe that can be clarify is it doesn't seemed that the population that surveyed is patients that are within the hospice of palliative care but it looks like it was every patients that expired. And I bet – I just want a clarification on that.

Mary Ersek: This is Mary. That is accurate. There are palliative care change most of the VA hospitals. But what that enables us to do is and we can provide this data. They've both been published in our part of our regular reports with the field. But what we're able to identify is that in facilities where you know, if a (dissidents) received palliative care, if he die in a hospice unit, if they've received chaplains visits. So these are a quality of care indicators or a process measures if you will.

We're able to see an associated increase in BFS score and that's on the performance measure that's on the one global item. And so then we're able to provide this information to facilities and also on a regional level which we call (inaudible) and the VA and to say, you know, you have for example lower rates. Your rates of palliative care, consults have dropped perhaps you know, the nurse practitioner left the team, or physician left the team and your scores went down. We know that you know, high increase or receipt of a palliative care consult is associated with better outcomes.

So that's how we do it. So you are correct. It is not limited people who only either die in a hospice bed because we have inpatient hospice units of the VA, throughout the VA or the receipt of a palliative care consult. Everyone is included.

James Merlino: Yes, and this is Jim and I think that's the appropriate way to look at this right? This is a global survey that looks (inaudible) …

Mary Ersek: Correct.

James Merlino: … life in measures that are across all types of interventions with the exception of some of the exclusions which they label like suicides and things like that that are (unaccount) – that you really can't control. And I think that in other way to frame this early discussion about the measure and the data or the ledger is that really, this is about it's a process measure that will allow us to improve the care, right?

The ledger supports that we have processes like this in place, care is improved. And this becomes a tool to allow VA administrators really to use the data to improve their process around standard of care. That's kind of how I thought about it.

Mary Ersek: And we can also examine a subgroup of people who got palliative care and able to look at, OK, when you do provide palliative care, we're now beginning to compile data about specific characteristics of that palliative care consults that will affect outcomes.

James Merlino: Exactly.

Len Parisi: This Len. When – I'm sorry. Just because there're enough names out there, when you make a comment, could you remind me who you're with? I had a question that related the last couple comments. And could you just clarify Mr. Brian Lindberg with the Consumer Coalition among the committee. Clarify whether we're talking about that there is a process measure already there and then now in addition, we're getting feedback from this tool to find out what the family were patient perceptions of that was?

Karen Pace: So, this is Karen Pace. What we just clarify a couple of things in terms of our, you know, criterion and how we need to look at this. This is a aggregated information for a facility about their patients global rating of the hospice care. So, it's not using all of the data from the survey. It's using the global item that asked overall how we do rate the care that you received in the last month of life. That's what they are proposing as their performance measure.

And in terms of the evidence, I think as Len said, you can certainly note that these evidence-based processes will actually affect this. Basically, that what

has been provided in the form, it doesn't say that. So, I appreciate you all filling in kind of your own experience in knowing that those things exists. Typically, we would want them to actually say what those are. So, providing chaplain services, providing prompts and effective pain managements, and other symptom management. It's not a lot that they had to say. But I think that you're kind of filling in the blanks.

In terms of the question about process measures, we don't necessary -- NQF, for example does not necessarily have process measures. Perhaps the VA is looking at those things. And actually part of the survey is measuring some of those things but it's not being proposed as performance measures for us.

So, what you're basically going to be judging is a performance measure that is based on this global rating and will have more discussion about that as we go along.

Brian Lindberg: OK. This is Brian. I just wanted to clarify. What was getting at was I mean you touched on a little bit. But what I was getting at is there already a process measure the VA is using related to whether and individual family had the consult because this isn't a process measure. This is – as you say, the survey and – so I'm just – I was just trying to link the two pieces up to see if there was it.

Karen Pace: Yes. No, I thought what you were asking is, you know, is there evidence of, if you will (actionable) processes that will affect the score. You don't use – well actually, we do use the processes of outcomes in a sense. But they are always linked to BFS. Is that – and to all reports, our focused on the global item. We also give information within the VA about each item. But we're not proposing that the process outcome. Does that answer your questions?

Brian Lindberg: I'm not sure really. So I guess if I were – and maybe this is just getting us off track. But if I were the VA, I would be just – let's just talk about whether someone gets an antibiotic, is it a right time before surgery? Is somebody getting this consultation? I would be measuring that separate from my survey that we're talking about …

Karen Pace: Correct.

Brian Lindberg: … today. And I was just trying to figure if you're already – if you had measure that separately as a process measure as opposed to a survey?

Karen Pace: Correct. Yes, we derive all our process measures directly from the electronic medical record. I thought that the original question was to say, you know, so this outcome measure actually related to certain processes that we would – we had to know from the literature. We would want to test have an effect on outcomes.

Brian Lindberg: Great. Thank.

Karen Pace: Right so, basically, I mean I think that answered your question. But it really isn't necessary for what we're asking here. The question really is, is there information from the literature study that there are processes and interventions that would affect the experienced that's being measured.

So, I think we'll keep moving on and let …

Mary Ersek: May I just – I'm sorry. This is Mary Ersek from the PROMISE center.

Karen Pace: Yes.

Mary Ersek: And because we're unfamiliar, it sounds as though you wanted us to provide more information. We have lots of book published and unpublished data. It's the part of the process to clarify. Can we submit that or that's what the purpose of this call is?

Karen Pace: Well, Lauralei had sent all the developers some instructions and examples for filling out these forms. So, perhaps, you know, we can go have and have discussions with you about that.

Mary Ersek: OK.

Karen Pace: So, let's move on to the committee members. Anything else you want to say about whether you – about whether there are – I think you're in agreement that there are structures and processes that an interventions that affect – will affect this outcome. Are the (inaudible).

Male: Yes.

Male: Yes.

Lauralei Dorian: Yes.

Karen Pace: All right, so Lauralei, we need to go for the slides. I guess this has been worked very well.

OK, we're going to go onto the next criterion which is 1b, Performance Gap and Lauralei?

Lauralei Dorian: So, performance gap asks whether the data provided demonstrates the gap and care which means the variability or less than optimal care in one setting or another that warrant a National Performance Measure. And we're also asking whether the submitted measure provided that demonstrate the disparity for certain populations (about) groups.

Carin van Zyl: So this is Carin van Zyl. When I went through this worksheet, I do think that the data that the measure developer's provided does demonstrate wide variation and the quality of care between VAs. So there is a performance gap there. They also submitted data that said there were racial disparities that needed addressing as well.

Is everybody else agreed?

Male: Yes.

Carin van Zyl: I agree.

Male: Yes.

Karen Pace: So, we're going to call this up from the submission form here. There we go. OK and so if you're in these measures submission form were at 1b.2 where we ask them to provide performance scores on the measures that's specify. And there's actually no data here in terms of the facility level scores that would be

constructive from this measure. Though, perhaps it's been another sections we need to look at.

Carin van Zyl: There's a whole other evidence section that was able to pull from the website which was still with lots of steps but I didn't understand. But that's where I pulled my answer those questions.

Karen Pace: From the NQF website or from …

Carin van Zyl: Yes.

Karen Pace: OK.

Male: Yes, there – yes.

Carin van Zyl: And it only – I'm looking at it right now but it was …

Karen Pace: It might have been …

Carin van Zyl: It looks a little bit like the website you try to pull up like way down but it was – where all the validity testing and the reliability testing.

Karen Pace: OK. Right, OK so …

Carin van Zyl: And that's where they put in information about gaps between (VSNs) the VA centers, right.

Karen Pace: Right. Right.

Carin van Zyl: Right. I didn't find it in that, the top (inaudible).

Karen Pace: With the next testing – the testing attachment.

Carin van Zyl: Yes. I think so.

Karen Pace: OK.

Carin van Zyl: I wasn't sure if I was supposed to look at that data as well.

Karen Pace: You actually look for everything. We ask people to put it where you're going to be asking the question because it makes your evaluation much easier. And we can follow up the measure testings – so people at measure testing its 2b5, identification that's specifically significant and meaningful differences. So they do come …

Female: Yes. OK.

Karen Pace: Yes, 2b5, page 8. So, yes, and 2b5.3 they said that the facility level scores showed considerable variation. But they didn't give us a lot of information. Normally, we like to see kind of more distribution information. It looks like they said, the unweighted range which would have been kind of including the very lowest to very highest is 33 to 95. And that the interquartile range which would be from the 25th percentile to the 75th was 53 to 64. So the total range is often, you know, that includes any outliers, et cetera. So it's just a little hard to tell from that. But the interquartile range gives you a good idea.

Carin van Zyl: So, does that qualify as an off data or is that where the NQF committee would say you should adjust that we need more information on the specifics before we would say that's about updated to support wide variability which was claim made somewhere else in the evidence submission.

Karen Pace: Right. And you're right. They did provide information about the disparities. And again, there's no kind of sets criterion for this. But the idea is that we have an idea of how this actual measure is about you know, is demonstrating variability and room for improvements and you all, we have some perspective on that or want to ask the developer a questions.

You know, certainly you know, the biggest chunk of facility is between 53 and 64. And perhaps, maybe we can ask the VA to just make a comment on that in terms of the spread of this course across their facilities.

James Merlino: But I would argue that their reliability, those pretty – I mean VA patients are very homogeneous. I think I would be more concern about this if you're trying to apply it outside of different healthcare systems. But within a single healthcare system that treats a relatively homogeneous population, I think it's usually pretty good.

Karen Pace: Who is that speaking?

James Merlino: Jim Merlino, sorry.

Karen Pace: OK Jim. And yes, we'll definitely come back to the reliability testing. Right now, the question is about the performance gap and I'll just pass it through if there's any – do you think that that range as an interquartile range signifies you know, that there are – there is room for improvement. That's the question.

Lauralei Dorian: If the developer want to comment on the different scores between facility?

Brian Lindberg: Yes, I mean I don't – I'm not concern about it again because of the population.

Karen Pace: OK.

Carin van Zyl: The other gaps that were identified were racial disparities but also disparities in the care provided so that you didn't have involved family members, correct? So, again we're not being an expert in the statistics. I think that there was at least reasonable data from me to say, OK, this sort of addresses performance gaps and a multiple dimensions.

Karen Pace: OK. That's good. So let's just keep …

Mary Ersek: Again the only – this is Mary from the PROMISE Center. I'm just going to say, have been able to you know, move that interquartile range of years. The other thing is there is a wide variation because again, some of our facilities depending on the quarter have a few debts. So actually – I hope it does and then fewer survey results.

So we have been able to move the needle if you will. So I think it does suggest enough variability and room for improvements.

Karen Pace: OK.

Carin van Zyl: So for my education, how do we decide that the data prevented is enough? Do we have threshold for the quality of data presented?

Karen Pace: There're no specific thresholds. So you know, the committee just need to (wait), you know, does they present the information on a large enough sample to give you an idea, you know, whether there – in the variability can either be across facilities or within population subgroups, the disparities issues. And this is really an area, the first thing is to be even have any data for you to even look at and then the second is it really does require some of your judgment and expertise, or thoughts whether – what they've provided is demonstrates that there's still room for improvement here.

Carin van Zyl: OK thanks.

Karen Pace: Great. So I guess that we could bring the slides back up. We can move –I'm just keeping an eye on the sign. (It doesn't) really do want to get the scientific acceptability, the validity and reliability testing. But in terms of high priority, we're actually for PROs, we're asking how it was determined, how – that the target population value of the PRO and finds it meaningful. So, we're asking whether there's any data submitted that the – a group of patient for example had inputted the questions that are asked.

I would like to have your – talks about that criterions.

Lauralei Dorian: So we need to go to the measures submission form and that would be in 1c.5 and it looks like that was left blank. So maybe we can ask the VA folks when they were developing their instruments, did you involve patients and families?

(Mary Ersek): Yes, we did. And this is a national performance measure at the VA. And (Jean) maybe able to help me with the numbers but we have 21 regional networks in the VA and I know that it's a little dangerous to talk about metrics in the VA these days. But it is used as a performance measure this in leadership and you've always had – I want to say, it wasn't between three and eight or seven. It isn't about a 21 that have chosen this measure. We had it chosen for them as a (performed). It's one of two (eccentric) measures in the VA. The other one of course being HCAPHS.

Karen Pace: So, can you tell us a little bit about how – and not in any great detail but just how you involved the patients and families as you were developing these PRO?

(Mary Ersek): I will probably call upon (Jean Lou) who work with the developer David Casarett in the (trauma) call and …

Female: Yes. Can you clarify, are you talking about when it was develop as a measure or when we develop it as a survey?

Karen Pace: Well, probably starting with the survey because that's where you identified what was important to patients and families that's a survey level.

Female: Right so you didn't clarify also – we started of with open ended questions and focus groups with family members.

Karen Pace: OK.

Female: And then from that we identified themes and different domains of care that families found important.

Karen Pace: OK. Any questions from the committee about priority or …

(Keith): This is (Keith) and I am the patient adviser on the committee. I'd like for the developers to think a bit on how patient's preferences are incorporated.

Lauralei Dorian: Well, the wording – wording of the questions is did – on some of the items is did the veteran receive the care that they wanted. There's one item that addresses just an overall care in general and it is that did they get what they wanted. And then there's also reference to you know, did they get the spiritual care they wanted. And so it's framed in not just the spiritual care but did you get what you wanted. And in some cases, no spiritual care is what the veteran …

Female: Right. OK.

Female: OK.

Brian Lindberg: This is Brian and I'm, you know, I'm not sure exactly how it fit in but it seemed to me that I – as I look at some of the data that given that out of the total populations that was considered potentially for this survey and that you

eliminate some right of the bat because the veteran guide very quickly after entering the facility or whatever. Then it seemed to me that there was still a large proportion of certainly, you know, just in straight numbers more than half. But if you – but percentage wise, more than that that would – that were interested in completing the survey.

To me that gave me some sense that this has value to the family. Now, some people just do it because they think they should. But to me I thought that the fact that they were willing to complete a survey at a very difficult time was a signal that this was a value to those people.

(Mary Ersek): And I agree. And if I can just say, you know, the survey itself will – now we (concept) that out. Imagine getting a phone call. We're calling from Philadelphia. You know, we're calling people all over the country including Puerto Rico. And so, that, one of the strengths is, it's not the people who'd care of you, or calling you. People have to feel as though you need to be nice or say the things. But on the other hand, the survey comes from a group that you've not heard of.

Yes, we obviously mail it out with the VA or probably address on it. But – and so I just want to add that as evidence that family is considered this important. We also – we get fewer handwritten notes that we did the phones (tray) but we still get those. We also get – I get several pieces of mail every quarter where one of us does where there is a long letter written back. And mind you, they never met me. I am – a name on a letter that comes from them and yet they write, you know, letters (back) experience.

Lauralei Dorian: OK, thank you. Now we will move onto the next criteria which is all about measure specification. And so, here we ask for reliability and validity. First we ask in 2a1 which specifications are unclear then we would ask that (PO) and specifically for – go back – slide up.

For a PRO PM, is it clear which questions or items in response options are uses for the performance measures. So we did a specific that the developers provided all of the questions and possible responses and are there survey sampling instructions for those who will be implementing the measure. And

do you have any concerns about the likelihood that this measure can be implemented consistently.

So I'll pause that and see if there are any reflections about that.

Karen Pace: So, is everyone clear on how the performance measure is computed?

Female: Yes.

Len Parisi: Yes.

Karen Pace: OK.

Len Parisi: I did one have one question that was related to a previous comment about the methodology for surveying (inaudible). The family care giver or the bereaved. How was the decision made whether or not it's a phone call or a mailed survey? That was still unclear to me. It seemed like from what I read, it was the first encounter. If there was an encounter by telephone then they would get a survey. So that's still unclear for me.

Female: Now so they get a survey and if they have not responded, about four to six weeks after the survey is sent out, we do a follow up call to one confirm that the survey is done. And then, I'm sorry. That they receive a survey and if they haven't receive it, whether or not they like us to resend it or they like to complete it over the phone. And within the mail – the initial mailing, they also gives us phone number for those who just – would prefer to do this – over the phone.

(Mary Ersek): I might want to add that this is a change earlier when we originally were approve or endorsed rather. It was predominantly a phone survey and we're not able financially to continue that predominant mode of administration. However, we didn't – people who didn't want to respond to a written survey either literacy issues of whatever.

So even though, it's not a predominantly mailed survey, people have the option to complete it either by phone and they do. You know, they get the

survey but they also have a phone number that they can call. So now it's predominantly mailed, but with a choice.

Len Parisi: I see, so as a mailed survey and then during your follow up calls, if they opt to respond that way. Is there a mechanism in place to avoid duplication?

(Mary Ersek): I believe and (Jean), you may want to – I believe it's whatever comes in first. We do have a set. You know, we don't decide, oh this is the highest score. We're going to keep this. We have a policy emplace that states the order because occasionally, we'll get two from different people in case we have to go to the secondary person. (Jean), do you want clarify how we make that decision?

(Jean): Correct. Just the one that comes in.

Len Parisi: And one last question because I know there are high response rates depending on the type of survey. Do you do any analysis on the response rates through the telephone versus the mailed in survey?

(Mary Ersek): Yes. We've already published the numbers font size for the phone survey. We're in the process actually of preparing a manuscript. We've conducted analysis on only the mailed survey. We – our response rate tab dropped and we've not yet – we need to talk with leadership as well as other people in VA quality measures to figure out what we'd actually do on non-response bias.

We have those data but we have – there's actually very fresh. And I can forward those but they were not – he did not have those (inaudible) when we initially submitted these applications. We have it on the phone survey.

Len Parisi: Great. Thanks. And since this is a new process for many of us, myself included, I have just one more question if you can take a minute to answer. Is there a particular differentiator between this survey and (fax) which was the motivation for creating this one?

(Mary Ersek): I'm sorry, a different?

Len Parisi: Was there a different purpose in developing your own survey rather than just implementing another tool that – such as (text)?

(Mary Ersek): You know, I would – I think this is all of them and Joan Teno maybe able to comment for I know that (inaudible). You know, we did want to make something that was specific to the VA. We – if you'll notice, we have a couple of questions that's on VA specifics benefit.

Len Parisi: Right.

(Mary Ersek): Although our performance measures actually, it's very close to the fact. I mean obviously we don't frame our question, you know, how you would you assess hospice care because not everyone got hospice care.

Len Parisi: Right, right.

(Mary Ersek): For all the questions that's (simple). And so, I have no doubt that we borrowed, there was borrowing all around, sharing I think give you the correct word. But we felt like there were certain issues. For example, we have a question about PTSD which may or may not appropriate by the VA.

Len Parisi: Great. Thank you so much.

Lauralei Dorian: Great. Great questions. And I think now, we'll just briefly touch before we get into testing upon specifications in terms of whether they're consistent with the evidence. So we're just wondering if you think the – after having reviewed the evidence or the specifications for the measure inconsistent with the evidence and is it consistent specifically for PRO PM, with the target population values and find meaningful.

Karen Pace: This is probably not a big issue for this particular measure or even the set of measures but will just be if there was any concern – concerns or issues otherwise we can move on catch things up.

Len Parisi: Yes.

Karen Pace: OK.

Lauralei Dorian: OK. Now we get into testing and this is where you'll open that other – the first attachment which is by acceptability testing attachments. And first, they look at reliability and just as a reminder for performance for PRO PMs reliability needs to be demonstrated for the patient level status so in other words the scale and after the computed performance scores.

So some examples of the performance measure testing might be signal to noise, (inter)-unit reliability or ICC and the Chrome Box Alpha is often what you'll see for testing of the scale itself.

So our first question is 2a2 which is, what level was tested, in other words patient level data as I just said. We need to have both patient level performance level data. And so we're wondering if testing was conducted with post levels and what the – testing with an adequate scope to generalize for (live track) implementation and what's an appropriate method. And within results describes and demonstrated (efficient) reliability. So I know this might be getting into territory that some of you are a little bit unfamiliar with and we completely understand that. So, I'll sort of open it up for either feedback on that question or questions about reliability testing in general.

Carin van Zyl: This is Carin van Zyl from City of Hope. When I went through this, I only found testing up a – the data level not at the measuring level. Did I get that right?

Karen Pace: That's kind of what we saw as well. So, we can maybe ask the developer and do you want to first anything about the reliability of the kind of data level or instrument level?

Carin van Zyl: I could but I don't feel overly qualified to have a very sophisticated view of that. That's something I'm trying to teach myself as we go along.

Karen Pace: Right, sounds great. And so, you know, they did the Chrome Box Alpha in terms of you know, looking at the internal consistency reliability of the instruments. And their results were 0.70 reliability from Chrome Box Alpha which is you know, acceptable results. And you know, if anyone else wants to comment on that, then we'll ask the developers specifically about reliability testing of that computed facility or based in level score.

Anyone from the VA want to tell us whether you've done any reliability for the computed scores that you're using for the facility level?

Mary Ersek: I guess I would just add some clarification on that. You mean Chrome Box Alpha is on individual?

Karen Pace: No. That the idea with the reliability testing is the level of the computed score is you're using this to come up with a facility level score which is – let's see. Is it a percentage of …

Mary Ersek: Family members who answer the global question, excellent versus all other responses?

Karen Pace: Right. So each facility will have a percentage score is what you're saying that the performance score is.

Mary Ersek: Correct.

Karen Pace: So the question is, when you look at those facility level scores, have you looked at – have you done any reliability testings.

Some examples that we've seen specifically with these surveys have been into unit reliability based on the (Nova) testing.

We've seen – go ahead.

Mary Ersek: And so, we do examine the scores across different types of units and we've conducted test and the – not surprisingly half of these units always or consistently – not always, consistently score higher than other units.

Our CLCs consistently score next with acute (failure) units followed by intensive care units.

Female: Right. And that's not exactly what we're looking for with the reliability testing. I mean we can…

(Crosstalk)

Female: Mary?

Female: More validity, right.

Joan Teno: Yes. Hi, it's Joan Teno. Maybe I could help Mary out a little bit. So looking for an ICC done at the facility levels. So, you would do it at your – from VA facility level and do it on ICC within (Nova).

Karen Pace: OK. No, we haven't.

Female: Is that something that you have dated that you could do?

Karen Pace: Yes.

Female: OK. Because it is a requirement for these types of measures based on, you know, the updated that, you know, we had a project a couple of years ago that really look at PRO performance measures so we – obviously, NQF is endorsing that performance measure. That's why we like to see that. So we can follow up with you after this call.

Karen Pace: OK.

Carin van Zyl: This is Carin van Zyl from City of Hope. Are there resources to the committee members that are easily accessible to help us be a little bit more sophisticated making judgments about whether it's statistical method, whereas the right one for the question that they are asking?

Karen Pace: We have some things. I'm not sure how useful they will but let us see what we might be able to provide to you that would be perhaps of some use here.

Carin van Zyl: Because at this point, my level of sophistication I said is that you tested at the data levels but not those and I don't feel qualified to say, well, because this circle must be (inaudible) that you picked wasn't the right one for the right question.

And that was such an expectation is all about, or in which case I should get some steel building done or whether we rely on the diversity of the member that somebody in the group will be able to comment on that.

Karen Pace: Right. It's an excellent question and be may not have succeeded in getting the diversity in each of the small workgroups we have tried and then had to move some people around because the availability of time. But that's something that we can try to do and I don't know if there's anyone else on the call here that once they make any comments about it or have some experience with the Chrome Box Alpha or the ICC, they've only provided the Chrome Box Alpha here, but if you have any comments or – OK.

There's something that we'll try to (fix), definitely all together for you or one of the things that we are considering in terms of the in-person meeting is having more information for you like that available that the staff can pull together so that …

Carin van Zyl: Even a glossary would be helpful.

Female: Right.

Female: OK.

Carin van Zyl: Because I never heard of a Chrome Box. I'm not sure what is best suited to do so I have no way of knowing whether that's the right thing to answer the question that we're asking us on.

Karen Pace: Right, OK. That's a good point. You know, then in the form, we do ask the developers to try to explain what they're testing but that doesn't always come through. But we'll see what we can pull together. And now, we have some things that we use for the measure of testing report that at least might be at starting point. So, we'll take a look at that.

Carin van Zyl: Appreciate it.

Karen Pace: OK. So, why don't we talk about validity testing then? So, for validity testing, similarly we want to make sure that it was tested, it was conducted at both levels and within appropriate method and we want to know what's the results were and gather the results to try how the measure either does or does not demonstrate validity. So that the conclusion about quality can be made

are valid. So, we might open it up if there're any thoughts or comments about validity testing.

Carin van Zyl: Well I think the number of entities in patients are always huge in VA study since that there's very little to criticize on that stand point but I will just – I know I have the same comment as I did for the reliability questions.

Karen Pace: Right. So, it looks like they basically did this again at the level – the patient level, the patient or level versus the facility level score. And I would say that some of the things I put under systematic assessment of phase validity was actually empirical validity testing.

So, they look at this correlation of this overall score at the patient level with things like patients who had a palliatives of care consult. Patients who died in hospice unit, chaplain con – and basically with most validity testing, it comes down to end of testing conceptual relationships that you would expect satisfaction or the total experience reading to the higher, you know, they may have hypothesized that. It would be higher for those who have a palliative care consults or that those who are actually in hospice unit may had a better experience or patients who have a chaplain concept is. And that's t basically what they did and they – and that would be appropriate testing both at the patient levels – patient score level and also that kind of testing would be appropriate at the facility level score though it needs to be constructed a little bit differently. But let's see if the committee has any questions, thoughts.

I think it was Jim you had earlier on say some observation about the reliability. Is that anything that you want to add about the reliability or validity testing?

James Merlino: No, I – I mean I'm not an expert in Chrome Box but I have some experience with it. I thought their metrics were pretty good but I – and I do understand the issue raised about inter-center in your inter-center variability. I'm not real clear on again whether the fact that it's a homogenous patient population of veterans really makes set a big issue here, but I will be interested to hear with the – about how the VA responds to that.

Female: Well, now they are mostly man. I would say that we do have a fairly broad. It is a national sample and so we are able to identify certain patient's characteristic as well as the – or the level characteristics. I mean for example, we know – we've documented that non-whites and older or younger patients have lower scores. We know that there are variations among facilities of VA facilities are labeled or categorized one through three with one thing the most complex usually urban settings, tertiary care settings, those typically have lower scores than the worlds smallest facilities do.

So, I do think that we have a fair amount of diversity and we're able to demonstrate that we can identify groups that are quite frankly at risk for lower or worse outcomes on the BFS than others.

(Jim Merlino): Yes …

(Crosstalk)

(Jim Merlino): Sorry, go ahead.

Debra Saliba: This is Deb Saliba. I just wanted to jump in as well about the VA. You may see within facility variation because there is significant variation across units being on, you know, who's operating and managing, what kinds of clinicians are operating, and managing particular units. So, you may see a big difference between CLC care, for example, and what happens within – on the general medicine unit.

So, the question may be, are those relationships in the direct (spec) more sort of a, that kind of validity as opposed to are they the same across the various units.

Female: Yes. And I'm not sure you're on the phone earlier, but my (talk) about a recent analysis that we did that is now under review, but looked at CLC specifically, but in – these are consistent findings. We've reported these, you know, for QI efforts as well as for secondary analyses.

So, not surprisingly when you look across all units. We're not able to identify the specific units that a patient died on, but we are able to categorize hospice

care beds, CLC beds which are our nursing home units, acute care and ICU. And that order that I just said those is generally highest to lowest BFS overall score.

Now, there are variations with our – or higher than some CLCs or even I suppose we can probably come up with an example of one that scored higher than the hospice unit. But in general, they are as expected.

Debra Saliba: Yes, is that variability within facility within facility or is that looking across all facilities?

Karen Pace: I believe that's also looking within facility, but I'm going to let (Dawn) because we have regular QI calls that where we …

Debra Saliba: We just wanted to give the members of the committee, you know, sort of a heads up, but given that finding that you've had that, you know, they may not have – that may not be the metric to hang their hats on in terms of looking – it may actually be more evidence of the strength of the measure as opposed to evidence against this performance.

Karen Pace: What metric are you talking about, Debra?

Debra Saliba: The (intra)-facility reliability.

Karen Pace: We're looking at interfacility. It's what – the point – and let me just clarify, the reason for that criterion is that, you know, reliability relates to measurement error. And the idea is, you know, performed, you know, in an accountability framework which is what NQF endorses performance measures for – as well as the improvement aspect is that you want to be able to distinguish performance so that, you know, you're actually measuring true difference versus a lot of measurement errors.

So, that's the basic foundation, but we can continue to have – we can – we'll work with the VA in terms of getting the additional information for everyone before the committee meeting.

So, maybe we should just move – maybe we need to move on.

Lauralei Dorian: Yes. OK. Just looking at the time, we do have a number of other questions about testing and then our usability and feasibility criteria. But you want to make sure that the other measures get adequate time from you as well. So, maybe we'll return to that if we have time, we'll certainly discuss all of those and you'll be voting on them during our in-person meeting.

But does anybody have any point that hasn't already been raised that they'd like to raise now all about – the rest of the testing questions were about exclusions and case mix adjustment and then our usability and feasibility about how the measure has been used and the feasibility with which it can be implemented. So, if there are any comments about that. Also on the score.

James Merlino: No.

Lauralei Dorian: OK, great. So, our next measure is 0208, I believe. No, sorry, 1632 which is the CARE, Consumer Assessment and Reports of End of Life measure. This is from Brown University at the Center for Gerontology and Health Care Research. And this is in mortality fallback survey that is administered to bereaved family of people 18 and older who died of a chronic progressive illness who have received services for at least 48 hours either from a home health agency, nursing home hospital or acute care hospital. The survey measures the perceptions of the quality of care either in terms of unmet needs, family reports, concerns with the quality of care, and the overall rating of the quality of care.

So, does anybody have any general comments about this submission before we get into the specific criteria?

Female: No.

Lauralei Dorian: OK. So, we'll begin as we did with the first measure by asking about evidence to support the measure focus. So, is the relationship between the measure outcomes, the PRO, and at least one healthcare action identified and supported by the stated rationale? And so I think I'll also throw in there just the performance which 1b, the performance data provided demonstrates that there's a gap in care, variability to warrant this national performance measure.

So, we now go to the evidence form. So, we'll bring up the evidence form now.

Does anybody have any initial thoughts about the evidence for this measure?

Karen Pace: So, it looks like again it wasn't exactly stated.

Female: Yes, I would agree, the evidence for the connection was not presented in the evidence forms.

Joan Teno: This is Joan Teno can I just quickly to respond to that?

Lauralei Dorian: Yes. Sure.

Joan Teno: Yes, sure. I think I sort of misinterpret what you want. What I can do is work with the NQF staff to rewrite some of these to present our findings to publish which note that regions have different levels of intensity of ICU utilization report, different patterns of unmet needs that – and constructions with quality of care and the rating of the quality of care that was published in (JAC). I mean, that would be one thing that we could do this.

I do have a question for the NQF staff. You know, we're proposing this as sort of mainly quality improvement and not for public reporting or accountability. Does that change the interpretation criteria at all?

Karen Pace: Well, it actually means we don't usually consider it for endorsement. NQF, one of the conditions for considering measures for endorsement is that it's intended for use for both improvement and accountability applications. And it's something that, I mean, we didn't get to use and usability would be the first measure we were talking about. But especially these measures that are coming back for endorsement maintenance, if they haven't been used in accountability application within three years, the question is why and I guess you're saying that it's really not even intended for use in an accountability application.

Joan Teno: No. It could be, you know, I think, you know, obviously part of where my effort in times has been going on is moving these working with staffs that ran

to move the family evaluation of hospice care and create the test processes while intensely working on this. This has had to do take a backseat measure in terms of my time and effort. So, while we help some ACOs to start using it, we have not had the opportunity to fully implement it from that standpoint.

So, you know, we are, you know, not there yet in terms of where we are with this measure because a lot of my time has been spent on making sure that we have a successful development and implementation of the test process instrument.

Karen Pace: Right – OK. So, I think you – I think this one also you want – at least this one also, at least didn't present testing – did you present performance?

Joan Teno: (Inaudible) testing at the individual level at this time on to go to or level that you would require where would require – now, would require further grant funding or measurement, can you do that or currently exploring that right now in, you know, that would be something for the future. So that, you know, since that measure legislation in response to the call for palliative care measures.

Karen Pace: Right.

Joan Teno: At that time, you know, that measure was committed to that call – it wasn't clear to me that was in three years, I had to go out and get, you know, 200 organizations using it.

Karen Pace: Well, it doesn't – there's not a certain number, and actually, you know, that's not a much stop criterion as some are, but there should be plans for what – how it's going to, you know, a reason why it hasn't been, and then a plan for how – what is the plan for to be used in the accountability application.

But I guess that the bigger question before we would even get to that is testing at the level on this performance for. I think, you know, your original question about do we look at this criteria differently if it's for QI only if that's the question that's been post to NQF whether they want to consider endorsement for measures of – at as QI only, but that's not our current process.

So, it is something that, you know, various task forces is looking at here at NQF, but our current process and, you know, the initial conditions when you – so under the sufficient form is that it's intended for both improvement and accountability applications, but why don't we see if the committee has any questions or observations about the measures, the information that's been submitted and we can, you know, perhaps some of follow up conversation with you offline in terms of how we should proceed with this, but so let me – committee members, you want to – how do you want to do this?

Lauralei Dorian: Yes, I get – maybe – let's just start with the evidence criteria, I know we went through each one individually for the first measure But, you know, now they know.

Karen Pace: She mentioned that is not in here but should we worked with us on it, so.

Joan Teno: Yes, I don't think that's a problem here at the article said that talks about the relationship of structure for an outcome or study in other part.

Karen Pace: Right. And just to be clear with everybody with outcomes in PRO, it's not that we're asking for systematic review of the evidence like we do for structure and process measures. What we want to know is that there is this found relationship between what healthcare entities can do to affect to be ultimate performance score.

And sometimes that based on evidence and hopefully most of the time it is, but we're not asking you do a systematic review, but first of all, to identify at least some of those things and then by the rational. So, it's not a huge amount of things that you have to do for that.

Before we move on to specification, maybe does the committee have any feedback other about evidence that, you know, or lack of evidence, your professional in here, you know, expertise? What do you think of this?

(Sherri): This is Sherri, I mean I obviously got there was some evidence in there. I don't know if was in other places in this submission, but I mean – I think it's, you know, a great assessment but and it's necessary because they're definitely is less of an optimal performance noted and patients don't, you know, they

have the untreated symptoms and the lack of the emotional support and they have talked about studies to support those, but if it does not meet the specifications of what we're looking for then, you know, that's a whole different story.

I guess my question is how much of these can just be common sense? You know, that we can all extrapolate relationships between what this could do for improvement and care. And for – I'm try to figure out the distinction between – did somebody spelled it out well enough or can we just imagine that of course there is a commonsense connection between people scoring better on this measure and the quality of the care that they received and conversely if you have that score. These are the things that you could do to make them better.

Karen Pace: Right. I think, you know, that's exactly right. That's exactly what we're asking the developers to do and certainly from your expertise, your kind of filling in the blanks that are there, but you're right, it doesn't have, you know, and this is for any outcome measure or PRO that and it comes in the NQF for consideration. It's really that there is something that healthcare can influence to identify at least some of those things and they provide the rational part.

But you're exactly right, it's conceptual generally grounded in some evidence but we don't require that there be an evidence review we could (inaudible) the developer to can and provide that conceptual relationship that there is something that healthcare can do and that will affect the outcome that's being measured. But you're definitely right with a lot of basis based on clinical experience and your knowledge of the field.

(Crosstalk)

Joan Teno: … spell it out? So, don't to know again – so, we provide you in the (context a listing) of structure process out terms in each of the measures that based on the NQF framework. What we don't provide you is actual correlation or evidence that shows you that if you do X and it will provide this. We can provide some of that evidence based on papers that was written on when I tell this question I put this on interpretation that you wanted to see a conceptual

model then is going to develop an instrument, I misread it, and it sounds like you of in additional conceptual model, you want with the actual evidence or either from the instrument, so or from other that shows that these relationships (inaudible).

Karen Pace: No, we're not asking for a review of the evidence and we can talk more about this offline. It's really with outcome and PROs that's kind of an exception to providing a systematic review of the evidence. And so, we're only asking for you to identify those relationship between what healthcare interventions and the PRO. But I think maybe what we're missing down is a – we're missing that attachment you'd talked about. Actually …

Joan Teno: Yes.

Karen Pace: … that …

Joan Teno: The article was …

Karen Pace: … the article got uploaded twice instead of – I think the appendix that …

Joan Teno: OK.

(Crosstalk)

Lauralei Dorian: I checked, we didn't receive that attachment.

Joan Teno: Yes.

Lauralei Dorian: So, if you'd sent that to us, we can make sure …

Joan Teno: Sure.

Lauralei Dorian: … the committee sees it.

Joan Teno: Yes.

(Crosstalk)

(Brian): This is (Brian). And I had a question in – at one point under the – for endorsement maintenance that says, "We report at the National Health and Aging Trends Study is using key items for the composite that it is allowed us to examine the change and the quality event of life care between 2000, 2011 and 2012. Could you say anything more about that?

Female: Sure. So we have a (inaudible) just about to be (inaudible) and basically shows the quality of care has diminished, there is higher rate of unmet needs for pain, management between 2000, 2012 that shows that at least family members who are on the community settings rate the quality of care is lower, it's dropped by 70 percent after adjusting for difference in age gender, place, (inaudible) they responded in.

So, you know, but unfortunately on the one thing I'm concerned about is on that we don't have a database where we've had on healthcare – enough of the healthcare entities calculate (IPC).

Lauralei Dorian: OK. Well, maybe if move on then specification, it could bring us – the measure specification. We can talk at the same time about 2A1 and 2B1. So that's reliability and validity of the concise specification. So that first one, we're asking other any unclear specification, you know exactly what the data elements are and the code in the sub steps in the calculation. And specifically so PRO (PM) that there which questions on response items are used for the performance measure and/or the sampling instructions clear for survey, instruction clear.

Do you have any concerns about the likelihood that the measure can be consistently implemented? And then in terms of specifications consistent with the evidence. We want to know if there are any inconsistencies and if there's consistent specifically we suppose to target population value and find meaningful. So, I'll pause there, see if there any thoughts about either of the topics.

James Merlino: This is Jim. This has to be an interview survey, I believe. And I don't – I find it that it would be very difficult to implement that consistently. I'm not sure if they reported that well.

(Joan): This is – you're absolutely correct as the initial instrument was defined, this is self-administered survey.

James Merlino: Yes, and it's also as I recall – some of my notes, and probably about – I thought it was a lot of questions were taken out, because it was like at an average 28 minutes long.

(Joan): That's correct.

James Merlino: I think that's going to be very – I think that makes it very difficult to implement. I'm not so sure that – it's practical.

Lauralei Dorian: Does the developer have any thoughts to that?

(Joan): No, obviously, you know, I think one other thing that you need to think about is how do you get from here to there? And I guess one other thing, you know, in terms of the natural life getting a measure to go from thing that was developed for research to getting it and use widely. Obviously, we would love to do some sort of work. But you have to realize that requires funding and money with an assistance to do that. And you know, it's possibly what we're heading towards is (inaudible) this measure.

But I think you need to think about how one get these measures and how one works in these areas that people really don't have a pressing need for meeting for, you know. Right now, probably ACOs and healthcare systems are more concerned about ACO cap than they are about dying patients. But, when you look at where they're going to be saving money? They're going to be saving money on dying patients. And if we don't have transparency and accountability measures, believed me, this is all going to blow up in our face very quickly.

Debra Saliba: (Joan), this is Debra Saliba. What were your response rates with the telephone survey?

(Joan): We had I believe somewhere around the 55 percent response rate.

Debra Saliba: About what you get with the mailed survey?

(Joan): Yes, and you should know that we've adapted this survey into that. And in fact we have done a (inaudible) check and we compared survey versus telephone administer. I think, you know, if we were, you know, as I would plan to in terms of developing this instrument further. We would obviously go and put this into a self-administered, you know, survey.

You know, we initially did that with in fact (inaudible) hospice care instruments, we had that down to self-administered survey. And obviously (inaudible) hospice is not being presented at this time. But that could majority of (inaudible) as a self-administered survey.

Debra Saliba: And one other question to – speaking to the link question. I mean I'm sure it probably takes just as long to do self-administer by telephone. But, that – for any of, you know, for any number of surveys. But, the question has – did there many –did you find that very many people terminated?

(Joan): No.

Debra Saliba: During the telephone interview or did most people complete?

(Joan): You know, what are fascinating is the assumption is that you can't talk to bereaved family members. And if anything bereaved family members want to talk about with their experiences.

Debra Saliba: Yes, that's my experience as well.

(Joan): Yes. You know, we tackle with 10 people as I said in my application. We'll have a, you know, some distress for doing the interview rather I'll echo the exact same thing experienced. I received letters from people saying thank you for the opportunity to tell my story that if one family doesn't have to go for what we went through. That would be a blessing.

(Sherri): This is (Sherri) and you know, having lived this very, very recently. I just find that really – I can even come up with the strong enough word to, you know, frightening, horrifying the reason why I couldn't go back in the hospital nursing. That the ACOs are, you know, focusing in something that's going to

prevent something like this term going forward and focusing on something like this. Because, this is critical and it will save money in the long run if that's what they're looking to do. Federal also improved the quality of care.

Male: No one disagrees with that. And believe me I've been through as well, on both sides. It is extremely important for reducing cost and improving quality. It's important, because it's the right thing to do for families. But, the issue here is not really leveraging our emotions. It's deciding whether or not, you know, we have an instrument that is kind to be applicable and easily administer.

Because, when you'll get the book of surveys and other instruments that we have to use to measure the patients. And patients families perception of their care, you know, they – the ones that are not – there's are not easy to administer, don't work very well. And so, that …

(Sherri): Right. And I agree with that. And I – the problem this also requires real person time as oppose to a mailed in.

Male: Exactly.

(Sherri): Staff time, staff hours.

Male: That I think is the barrier to feasibility here.

Karen Pace: OK. And maybe before we move on to the next measure, Joan, could you just explain how the weights were came up with in terms of this formula. Is this to compute the individual tool or can you just explain us this?

(Joan Teno): Sure, the weights are based on the factor analysis (inaudible) …

Karen Pace: OK.

(Joan Teno): … estimate.

Karen Pace: OK. All right. Thank you.

Lauralei Dorian: All right, is there any web learning questions or comments for 329? So, before we move on to this the last measure?

All right. Now, I'm done. Last measures 0208, the family evaluation of care in the national health palliative care organization, and the denominator represents the number of survey with responses for at least 14 of the 17 questions. And the numerator is the sum total of the weighted incidents of problem scores referring and response of17 specific items on each survey. And if that's the survey is an after death survey administered to the bereaved family caregivers and individuals who died while enrolled in hospice. And so, as with the previous two measures I'll open it up to see if there any comments about evidence to support in that measure focused …

Karen Pace: Let’s go to the evidence attachment.

Lauralei Dorian: We'll bring up the evidence attachment.

Len Parisi: So, Lauralei, this is Len Parisi, I have a question. So, I'm learning more cumulatively here – cumulatively as we go through this. So, what we're being asked to evaluate is the composite score for this as the measure. Since that – and with the intent that be a publicly reported measures, is that correct?

Lauralei Dorian: That's correct, yes.

Len Parisi: So, it's not the instrument as we know as specs but, also – but just that measures. OK. Thank you.

Lauralei Dorian: Right, so it's the measure that's looking at the facility the accountability and the performance of the facility itself as oppose to the patient level.

Karen Pace: Right, each patient when get us for on this. And then that's aggregated to get the facility level score.

Lauralei Dorian: But much of it, you’d assume that the evidence of the survey will be somewhat to the evidence (inaudible).

Karen Pace: Well, the evidence is related to the outcome that's being measured. So, in this case you could conceptualize that as they said incidents (and problems). Go down to their – what they provided.

So, evidence during the measure testing. So, in this measure testing quantity, opened it up, they did show that there's lots of different things that the healthcare intervention services relate to the composite score.

So, for example, medicine received for patient's pain, feeling with the grieving, help patients feelings of anxiety or sadness. And there's a long list here, which is fine. And, you know, they have a – they did list some of the references that support that.

But again, this is – what we're looking for, this is for the developer to identify what kind of intervention and services relate to the outcome that's being measured. And so, I know this is a little hard for all of us to get our minds around because these outcomes are experienced with care. But ultimately, I think by definition that's what healthcare services and intervention should be making an impact on the patients overall experience with care. But I'll stop there and see if one has any questions or comments about or questions for the developers.

Joan Teno: So, with this was Joan Teno, given that I – you have some involved with, can I ask our clarifying questions on whether we provide you with evidence. So, for example we have paper that was published that said, (inaudible) programs that provide a more direct patient care volunteer hours have a higher satisfaction, would that be a sufficient health (inaudible)?

Karen Pace: Thanks. So, at this level we're really – at this level we're actually looking at the patient level outcome and the things that healthcare services do to affect that patient outcome. So, that's probably supportive and you could list but, the real key is just to identify what it is or several things that healthcare providers do to that influence patients perceptions and experiences based on what being measured. So, generally, going to be reflective of the types of questions and domains that are in a particular experience with care survey.

Female: So then that would be what they have listed here,

Karen Pace: Right.

Female: Correct?

Karen Pace: Right.

Female: Nice, OK.

Karen Pace: Exactly. So, the questions for this committee is – does this make sense and, you know, did they provide – they obviously identify these things and these things that does make sense, is it probably support accordingly literature. They're providing some references in 12. And if you go down to 1a2.1. So, this is the type of information that we – we're asking for. And it's really up to the committee whether you have any concerns (inaudible).

Female: This seems to me to spell out the things that hospices could either improve their score. So, to me it's means, this is at least the way I read it. It looks like it answers the question about, these relationships between outcomes and things that hospices could change to make their scores (better).

Karen Pace: Right.

Len Parisi: So, the challenge on – this is Len Parisi. And I may have a lot of experience with the implementation of this. We use it on our hospice. And have used it for years just take in the composite score in order to make in meaningful information to make it – to implement any performance improvement initiative. You need to do more of an analysis of what contributed to that.

So, for example the measures that would roll into the composite score are really be important data points that you would want to evaluate to see where you can actually make a change in practice.

Karen Pace: Right.

Len Parisi: And I'm interested and understanding how the composite score would be able to give an organizations information to make those change of …

Karen Pace: So, that's a good point and it's really something that sums a lot with any outcome measures. So, whether it's experience with care outcomes such as this or for example mortality that these kinds of outcome give an indication of how those organizations performances in concurrent sends to others but improvement really does require organization to dig in to their data. And as you said that, you know, that they look at the specific items and performance they go into this to really figure out where they need to target their improvement effort.

But that doesn't necessarily detract from being outcome performance measure but it is, you know, definitely one other things that outcome measure is that. Outcome measures don't tell you what you have to do, it only to there an area that you have to see more investigation and figure. And it could be different when every facility or organization what they might have to do to improve.

Len Parisi: Great. Thank you. There's one other things that is a bit of a limitation from my own personal experience. And that is the lack of other languages. And I'm sure we're going to get to that at some point we go through.

Lauralei Dorian: OK. All right. So, are there any other comments about performance gap?

Karen Pace: Well, let's go to performance gap, did we get that information?

Lauralei Dorian: So, the performance gap inpatient things to be as the patient level but I think we can ask Carol clarify. Carol, can you – this one with 1B2.

Can you tell me if it's – the …

Carol Spence: Yes.

Lauralei Dorian: … performance is actually as the organization level.

Karen Pace: Right, for the data that, you provided in 1B2 is this about organization level scores or this patient level scores. Could you talk about 228,000 survey and average? So …

Male: Give me one second, I'm reviewing my work.

Karen Pace: I'm going to ask Matthew.

Female: Yes.

Male: So, on this specific question, these are at the patient level.

Karen Pace: OK.

Male: I will add to that though if you – look at the break down of organization level scores and the testing data …

Karen Pace: Right.

Male: … there's a great graph under two 2B5.2 …

Karen Pace: OK.

Male: … this really shows a fantastic distribution of performance scores across the organization level.

Karen Pace: OK, it's good. And actually, yes and I had to (bug) this, I believe you actually provide them by (inaudible). So, those on the Webinar can see this or?

Len Parisi: Yes.

Karen Pace: OK.

Lauralei Dorian: For those who are less familiar with (inaudible) result, are you able to walk us through what this mean?

Karen Pace: Yes.

Male: And I'm going to correct myself actually, the performance score is that the – on 1B2, that is at the hospice level.

Karen Pace: OK.

Male: I'm sorry. That is at the hospice level. You can actually see the change in performance score from 2011 to 2013. Those are all hospice averages.

Karen Pace: OK. Thank you. So Matthew, do you want to do just briefly kind of go through this table, the distribution just to help familiarize our committee.

Matthew Haskin: Sure. The table of the distribution, I'll just quickly describe it. What you're seeing is the average hospice levels scores. The reason that we take the component, not that kind of question to the individual response composite scores and average them up cost the organization.

So, what you're seeing in this table, the average hospice score on the X axis and the Y axis is a percentage of included hospices in the survey that have that score. Based on this graph, it's a nice histogram, you're seeing towards everywhere – anywhere between 0.757 to 0.9825 which to me looks like a bit of an outlier but that's a different story. But you can see a very nice, you know, standard normal distribution across the sample.

Is that – am I clear enough on that explanation, do you need further?

Male: No.

Karen Pace: Fine. OK, all right. So, we have information about performance gap which is useful. Any questions from the committee?

Len Parisi: Well, I would like to comment about the language, if it's not in multiple languages, it does not address the large population spending on the service area for the hospice. Is that in applying for the future or is that a requirement of probably reported measure or survey?

Karen Pace: It's not a requirement but I think it's a good question and probably comes up, you know, I'm not – I'm not so sure where that falls but I think it's a good question in terms of probably relates to what percentage of a population this kind of survey would actually cover. And therefore, you could conclude about the score. So, I think it might kind of getting to the right issue for you.

Len Parisi: Yes.

Karen Pace: So, basically like you're saying because it's only an English language, is that correct?

Len Parisi: I believe that was what indicated in here. I should – we should ask the developer. Is that accurate?

Female: We do have a Spanish version. And in fact, what we do is we provide it on request. What we have found however is depending on the area of the country, people – they users comeback to us and say well, you know, this particular awarding isn't consistent with the way our Spanish population uses, you know, terms, etcetera. So, what we then do is tell those users to go ahead and make whatever, you know, local colloquial changes they feel are necessarily.

We have – for that reason; we have not collected separate data on the Spanish version. So, we don't have, you know, data to and a repository like we do on the English version. But, you know, again upon request, we've been able to make – we've made the – that Spanish language version available.

Karen Pace: So, I think to that issue that – and it's a good point to bring out that the performance score would really be – needs to be characterized more that its performance with English speaking population.

Female: Not necessarily. I mean, if maybe English as their first, you know, may not be either their first language but, you know, hospices do, you know, we get the demographics that indicate that we definitely have responded to our Hispanic.

So I would not say that it's limited to an English-speaking population necessarily.

Karen Pace: OK, all right. Let's move to specifications committee.

Lauralei Dorian: Yes. So again, with the specification clear, do you understand or would you have any concerns about their ability to be (inaudible) reliably and also, did you see any evidence that the measured population find specifications meaningful and appropriate? And there clear survey instructions for how to implement the survey?

Karen Pace: Right. So, one of the questions, let's go to the (S18).

OK. Carol, maybe you can explain how this is computed because it's hard to – I think we would need more clear stratifications but maybe you could just explain that and then we could see if we need elaboration.

Carol Spence: Again, I'm going to ask Matthew do it because he's the one that actually does this computation, just to say though in number three there, that is our weighted formula for each of the – that apply to each of the questions from the survey that goes into the (set) composite, but Matt can give more details in that.

Matthew Haskins: And I'm going to have to ask for clarification. Something happened with my phone right in here, where exactly you're at. So, can you tell me where you're at again?

Lauralei Dorian: This is (S18) which is that calculation algorithm as your (logic) under the specification section?

Matthew Haskins: Measures specification, OK. Hold on a second.

Lauralei Dorian: So, this where you have the different steps starting with the same data.

Matthew Haskins: F – do you say, F?

Lauralei Dorian: Sorry, F as in (Fan).

Matthew Haskins: I'm looking in strange places. I'm very confused. OK. So, what we're doing here is where step one is obtaining the data and basically the hospices give us the data. We then take the data and we dichotomize the responses into either the most desirable response or all other responses. We exclude from these responses those answers which equate basically to no answer. So, we put – we set those slides as no.

And then, what we do is we take the dichotomized responses, and we plug them into that – the calculations you see under step three. So the composite score is that calculation and every time you see – for example, F1, F1 is the

dichotomized response for the question F1. This is done at the survey level and these create a specific composite score at the survey level. From that point, we take the total number of computed composite scores and we average them up to the organization level to make – create the measure scale.

Karen Pace: So, is F1 of question or is that a multi item like domain – is each of those just a single question F1, F2, F3 …

Matthew Haskins: Correct.

Karen Pace: OK. And do we – could you provide the details of F1, what's F2 …

Carol Spence: Yes. We have those listed. So there are 17 questions on the survey that feed into the composite and earlier on we have each of those questions with its own number and then the questions itself with it.

Karen Pace: Right.

Carol Spence: Higher up in the submissions.

Fe Karen Pace male: OK. And I see that we have a data code for the – OK.

Lauralei Dorian: That was included as an attachment.

Karen Pace: OK, all right. OK. So, any question from the committee about the specification for this performance measure?

Male: No.

Brian Lindberg: Hello. This is Brian. I just get a little confuse there on the – on number two there, A is most desirable response and then D is all other responses for the questions. So there isn't a, you know, the equivalent of one to five in terms of positive response or got desirable response.

Karen Pace: That's correct. Somebody – we don't have the same response set for each and everyone one of these questions. Some of them are Yes-No questions, others

are have, you know, three-response options in, you know, too much, just right, too little.

So, what we do is we're looking for the most desirable response to each one of these questions. That is what we want the hospices to achieve, so that's the goal. So then, any other respond is considered not desirable and it's a problem. But this is why we dichotomize them that way.

Brian Lindberg: Thank you.

Karen Pace: OK.

Lauralei Dorian: All right. Well I see we have just 10 minutes left and we do need time for public comments, so…

Female: We need to…

Lauralei Dorian: We touched upon – yes, testing before but are there any other (inaudible) testing attachments for this. The liability – I believe Carol you said that it was – the testing was done at both level, correct?

Carol Spence: That's correct.

Karen Pace: OK. And why don't se see if any of – the committee has any question as we're going through this, otherwise maybe will just ask Carol to briefly explain the testing. Committee, any questions? I know that (inaudible) some of the questions about the Chrome Box Alpha. And then this – here's an example of testing it at the level of the computed facility score.

Lauralei Dorian: So they will be able to see the differences on the strategy that was used.

Karen Pace: Right. So, Carol or Matthew, do you want to just briefly give a description of what each of these test says, the Chrome Box Alpha through level in the unit reliability.

Matthew Haskins: Sure. So, the Chrome Box Alpha also is really – it's testing the instrument reliability. We're looking at the reliability of the question at the individual scale. And what is really looking at is the internal consistency of the 17

questions which we used for the composite score. And the results of it are pretty clear that we have a very strong Chrome Box (Alpha).

It's really hit that – the Chrome Box (Alpha) score was 0.809, just about as good as we can expect for that internal consistency. And it's really just saying that those 17 questions are really measuring the same thing in this case. We're saying that the same thing is in perception of quality or hospice quality.

The computed performance measure scale, what we did, we did a single-to-noise analysis. Using analysis variance and computing an (FIE) basically similar to noise, tells us how much information the results are getting are really true differences in variance. So, when you look at the score, there's variance in the score. And the single is true variance of measuring differences between organizations. And the noise is just random other stuff that's coming in. And when we look at our signal-to-noise analysis, we see a 0.76 which are, yes, 0.76 which is also very good.

Karen Pace: OK. Any questions from the committee about that?

Female: Not for me.

Male: No.

Karen Pace: And they also did validity testing at both levels. I know we're rushing here, but maybe, Matthew, would you just quickly describe your approach to validity of the instrument or patient level score versus the computed performance score?

Matthew Haskins: Sure. So for the measure scale for the instrument level, what we're looking at is we compared rater scores to each of the – well, their composite score which just as a review is a measurement of the 17 component questions. And we compared it to what the respondent assessed the organization quality between excellent, very good, good and fair.

And we basically, we used some basic statistical techniques to see how well those scores related to each other with the assumption that if somebody is – if

a composite score is well-correlated through somebody's rating of excellence, we're going to – the measure is testing what we think it's testing.

For the computed performance score, we basically used the same excellent – I'm sorry, the same global measure in that that the individual's rating of their perception of quality. But, we just – we leveled it up to the organization level. So, for the computed performance measure, the organization level, what we did is we compared the averages of the composite scores of organization to the averages of the global rating score to determine how well-correlated they were.

Karen Pace: Thank you. And, let's go then to the results.

Matthew Haskins: So, what you're seeing in the results – for the measure scale, what you're seeing is the average composite score is this mean – in the mean column and kind of, unfortunately, in your version, it's across two pages. But, there you go, that's a good look.

So, this mean column, the second column or the third column over, that's the average composite score. And then you're seeing the excellent, very good, good, fair, and poor ratings based on the average – and what the average composite score was for individuals relating or rating that hospice quality is excellent, very good, good, fair and poor. And you can see a direct relationship where as the composite score goes, up the rating goes and improves.

For the computed performance measure scale, the correlation shows as we would expect a significant and direct association whereas the average number or average number of individuals rating the hospice as excellent increased, so did the composite score for that organization.

Lauralei Dorian: Are there any other last minute questions or thoughts about this measure?

Female: Not for me.

Male: No.

Female: (Inaudible) as I have to run, I have a family meeting.

Lauralei Dorian: OK, great. Well, before we end the call, I just like, (Amy), the operator, to open the line for public comments to see if we have any members of the public comment called.

Operator: At this time, if you would like to make a comment, please star one on your telephone keypad. We will pause for a moment to compile the roster. Again for public comments, please press star one.

Lauralei Dorian: OK, hearing no comments, so I would just like to thank all of you again for very a diligent work today. I know it's – (inaudible) more than two hours. But, we feel like you're a little bit more comfortable with this criteria and our process. As always we're here, if you wanted reach out with any other questions, we'll be happy to setup a call. Thank you to the developers as well for being on the call and being available through respond to questions.

Next steps so that we will summarize this workgroup meetings and send them out to you. As we know, we have our in-person meeting at the end of July. You should have received, either today or tomorrow, information from our meetings department about booking, travel and accommodation. So if you don't here from them by the end of tomorrow, please let me know and I can make sure they get in touch with you.

And as I mentioned, we will be assigning each person as a lead discussant or secondary discussant for in-person meeting for each measure. So just be on the lookout for that information.

Let's see if there are any questions about upcoming stuff?

Operator: There are no public comments.

Lauralei Dorian: OK, great. Well, thank you again and we look forward to seeing you in July. And feel free to call in for the (inaudible) to listen.

Male: Thank you.

Male: Thank you.

Female: Take care.

Male: Thanks.

Female: Thanks.

END