NATIONAL QUALITY FORUM

Moderator: Suzanne Theberge
September 11, 2012
3:00 p.m. ET

Karen Johnson: Welcome to the NQF Neurology Endorsement Maintenance Project Phase 2 and Workgroup 2 Dementia Diagnosis and Symptom Assessment. This is the NQF team and here at the NQF office we have Suzanne and (Katie), and I'm Karen and I'm going to hand it over now to Suzanne who is going to take a roll call for us.

Suzanne Theberge: OK, good afternoon everybody. I just want to run down the list of workgroup members and then developers, so we know who's on the line for our transcripts. Gwendolen Buhr, are you here?

Gwendolen Buhr: Yes.

Suzanne Theberge: Great, Jolynn Suko?

Jolynn Suko: Yes.

Suzanne Theberge: Salina Waddy?

Salina Waddy: Yes.

Suzanne Theberge: Tina Cronin? Jordan Eisenstock?

Jordan Eisenstock: (inaudible).

Suzanne Theberge: And Sam Fazio?

Sam Fazio: Yes.

Suzanne Theberge: OK great, thank you. And for our developers, do we have folks here from the Pharmacy Quality Alliance?

Julie Kuhle: Yes, Julie Kuhle.

Suzanne Theberge: OK great. And how about AMDA?

Female: Yes.

Suzanne Theberge: OK great and AMA-PCPI?

Diedra Joseph: Hi, this is Diedra Joseph. I'm joined by Mark Antman and Jimmy Drozda, and Dr. Jerry Johnson will be calling in as well.

Suzanne Theberge: Grea, thanks very much. OK, so it sounds like pretty much everybody is here and I'm going to turn the call back over to Karen.

(James Lett: OK, if I may, this is Jim Lett). I'm on the call with AMDA, also with Jackie Vance.

Suzanne Theberge: Oh great, OK thanks for letting us know.

Paul Katz: And this is Paul Katz, I'm also on the AMDA team.

Suzanne Theberge: Oh great, OK, thanks for letting us know. Anyone else?

Jerry Johnson: Yes, Jerry Johnson on the AMA-PCPI team.

Suzanne Theberge: OK, thank you.

Karen Johnson: OK, great, thank you. Before we get started, I do want to say welcome to Phase 2 of the project to Sam Fazio. He is a new member to our committee and he is our Alzheimer's expert, so welcome to our committee, Sam. We're glad that you could come.

Sam Fazio: Thanks.

Karen Johnson: Phase 2 of the project is quite a bit different in a couple of different ways from phase 1. So in phase 1, we did our stroke measures and all of those measures were tested for reliability and validity and you will have notice by now that many of the measures that are in phase 2 of the project have not been tested for reliability or validity.

So, Suzanne had sent an e-mail out explaining what's going on there, so hopefully that didn't confuse you so much but that is a little bit different. So for many of the measures, we won't have some of the 20 problems that we ran into in phase 1. But another difference is that – before we go on, can – can all of you guys mute your lines, if you don't mind because we're getting a feedback on the line here.

That's a little better, getting a little bit better but – and to mute it's star 6.

Suzanne Theberge: Star 6 should be mute.

Karen Johnson: OK. That – that was great.

Suzanne Theberge: Thank you.

Karen Johnson: Thank you.

The other thing that distinguishes phase 2 from phase 1 measure is – is really the evidence portion and for several of the measures the evidence – well, let me back up and say for many of the stroke measures that we encountered in phase 1, there was an abundance of – of evidence underlying the measures and oftentimes, it was really strong studies randomized control trial, et cetera and we won't always see that in – in this measures for phase 2.

So, I did want to just remind you in thinking about evidence particularly that what we're trying to get at with criteria number 1, importance to measure and report, what NQF is looking to endorse are measures that are most likely to drive improvement in health care quality. And we do want to recognize that there are many, many important things that should be done in clinical practice.

You know, just lots of things that are very important but they may not rise to the level of being a national standard for quality improvement measure because we realized that it takes resources to collect data, to do things like publicly report scores and that sort of thing. So again, the – the bar, if you will, for a national measure or national standard is quite high and that's why we ask that the evidence underlying the measures, we would like that to be as strong as possible.

And oftentimes, that is quite easy to do if the measures are – outcome measures or even process measures that are very closely linked to desired outcomes such as, you know, certain treatment protocols or – or something like that but sometimes the – the measures that are more distal to outcome often just do not have those – that evidence based underlying it and that is one reason that we have here the hierarchy of preference again for measures that we explained in the tutorial.

So, we again we would prefer, if possible, outcome measures and then if – if not those, an intermediate outcome measures and then if not those, then again process measures that are closely linked to outcome. And I think the other thing to point out without getting back into tutorial mode, is when you're thinking about evaluating measures on evidence, we – we do hope that the evidence underlying the measures is presented in such a way that quantity, quality and consistency of the body of evidence is – is quite transparent.

So that everybody understands what is out there and what is there and – and possibly even what is not there. And it is easiest, is – there are systematic reviews in that sort of thing that are done where the evidence has already been accumulated and – and even graded. That makes it very easy to see what's there. That's not always the case and I did just want to remind you that when grading systems had been used, we don't specify a particular system but we do ask that if systems other than the grade or the US Preventive Services Task Force Systems if use that the – that be described – the system be described.

I think the other thing to point out is sometimes at – there are things like clinical practice guidelines that are used to support measures and those are fine but I believe we all understand that not all guidelines are created equally, so not all of them are always

evidence-based. Many times they may be evidence-based but they are not particularly clear in their materials about what that evidence is.

So, it – it can be quite difficult for measure developers to – to supply that information and if possible that's when sometimes if – if things have been graded, if the evidence has been graded, even in the guidelines, that could give you a clue as to what the underlying evidence is. But the other thing to remember, there is – we do not consider expert opinion to be evidence.

We're looking for demonstrations of empirical evidence. So with that, I think I'm going to stop this quick tutorial and I think as we go through these measures, I want to just point out there is a section in the evidence section, it's 1c.1 and that is the structure process outcome relationship item and that is actually a very important item that hopefully the developers are able to fill out because what that does is that actually gives you an explicit statement often of the link that they see between their measure focus and a desired outcome.

And then once you know what that link is that tells you right there the kinds of evidence that you should be expecting to see described in the following sections. That link also gives you a flavor of the proximity of the measure focus to the desired outcomes. So, that is actually a very key piece of information that's very useful and I think the fundamental question when it comes to thinking about evidence as – as we looked at the phase 2 measures is, does this evidence meet NQF criteria for quantity, quality and consistency of the body of evidence.

So again, there are many things that are very important to do in practice but may not quite rise to the level that is needed for a national standard. So with that, I am going to get it started I think on measures. We have a very full slate today. We had six measures that we're going to be discussing and we have approximately 15 minutes each.

So, let's try to stay on target if we possibly can and what we will ask you to do, we have a discussion lead for each of the questions and lead folks, what I'll ask you to do is very briefly just describe the measure itself and then if you were part of the phase 1 project, you hopefully, will remember that we discuss things in terms of the criteria separately. So, we talked about impact and – and had discussion on that if necessary and then we went to performance gap, et cetera. So if we can pattern our call today in that way, that would be great. So with that, I'm going to turn it over to Gwen Buhr, who is going to tell us about measure 211.

Gwendolen Buhr: OK, so this is measure 211. It is an antipsychotic use in persons with dementia and its put forward by the Pharmacy Quality Alliance. This measure is the percent of individuals, 65 years of age and older with dementia who are receiving an antipsychotic medication without evidence of a psychotic disorder related condition and so they take from their administrative data, that's the right word, people who have an ICD-9 code diagnosis of dementia or people who have been receiving a dementia medication, any cholinesterase inhibitor or Namenda, and those are the – the numerator or the denominator, yes.

And in the numerator are people who have had at least one prescription and more than three-day supply for antipsychotic medication who don't have schizophrenia, bipolar disorder, Huntington disease, or Tourette syndrome. So, the impact there is – so not only – only two people from our group reviewed this measure on the summary statement or summary thing that we got and one said that had high-impact, one said that it had moderate impact.

So, I think that a lot of people have dementia and a lot of people with dementia have behaviors and received antipsychotic medications. I think there's a lot of data on that, so I would say that has a lot of impact and that the performance gap has been shown that people are receiving antipsychotic medications when they don't have a diagnosis of psychosis and it's aligning with the CMS initiative to decrease antipsychotics in the nursing home.

Karen Johnson: Gwen, let's go ahead and ask the committee or the workgroup members, did anybody had any difficulties I think with either the impact or the opportunity for improvement, or pretty convinced that – that both of those would be OK?

Female: I mean I think it's an important topic and there was clearly evidence that this was both needed, this was for performance gap and it could provide high impact.

Karen Johnson: OK, great why don't we go unto the evidence piece?

Gwendolen Buhr: OK, so the evidence – so what they're saying is that this is a process measure and that they're trying to say that the outcome is death because antipsychotics and use of antipsychotics in people with dementia, people have an increased risk of death and the evidence for that is meta-analysis from randomized control trial, so that's the evidence.

So, I think that there is in terms of quantity, quality, inconsistency, there is a lot of randomized control trials and they had been put together in meta-analysis. So, I think the quality of the evidence is good and it's generally in the same directions of consistency is there.

Karen Johnson: OK, great. I noticed that on page seven of the measure submission, the developers actually noted in terms of some contradictory evidence, some differing opinions about comorbid conditions. That is something that we should probably remember to discuss that there's a validity section of the measure.

Gwendolen Buhr: OK. The other thing right under that, that I thought was interesting and wasn't sure how to put that into this discussion was saying that there is some instances when behavioral disturbances or agitation in patients with dementia maybe treated and successfully with non-pharmacologic method.

Therefore, pharmacologic treatments with antipsychotics may be chosen and it's implying to me that those are appropriate but I don't know how that would fit in because this measure just measuring everybody who is on an antipsychotic and – and presumably saying that it's all inappropriate because you – I don't know.

Male: I think that would be an overstatement clearly. I think virtually all of the practice guidelines would suggest our circumstances in which a trial of an antipsychotic is warranted and it – it would be those instances in which there were behaviors that were life-threatening for the individual or for the people around him and there was no other intervention that was –.

Gwendolen Buhr: Exactly and so that's sort of what I was saying is that, that's the case but that's – this is sort of nuances and the way that I'm seeing this measure is that it's just measuring black and white and the goal of the measure is to have as low a number as possible. When there are instances when it would be appropriate to have antipsychotics in somebody with dementia.

Male: Right and there's certainly maybe settings in which you would find more of those patients and otherwise. I don't think we'd want to create a measure that – that resulted in a disincentive for facilities to take the most difficult patients because – because they're coming out of hospitals where antipsychotics were necessary in order to control dangerous behaviors. And I would worry about having difficulty finding a place to discharge those kinds of patients in the event that this measure were taken to the ultimate degree.

Gwendolen Buhr: Right.

Male: Saying – I mean I'm sorry, saying that I – I would agree that in general the lower – the lower this number, the better. I don't think there's any disputing that but the question is what is – is 0 really – would that really be the appropriate goal for every setting.

Karen Johnson: And this is Karen from NQF and let me just put a little bit of context around this section. I think the developers were easily could have put this under section 4c.1 which is, he talks about susceptibility to inaccuracy, errors, or unintended consequences. But again, what they're saying is this can happen but they're kind of assuming that's it kind of the uniform distribution across facilities or plans actually. So, that's their thinking that it wouldn't necessarily affect the validity of the measure, so at least that's how I read it.

Jerry Johnson: This is Jerry Johnson, I'm not on the workgroup but I don't know if I can comment on this or not. Is it OK to comment?

Karen Johnson: Yes, sure.

Jerry Johnson: No, no I -- my colleagues certainly take care of a lot of patients with dementia and everything that was said about the rationale for using as low – a small amount of antipsychotics and trying not to use them all, all that I agree with but I don't – I'm not hearing that this – this measure is a good measure of quality in the way that it's constructed for the reasons that was just stated by someone else there.

There are absolutely indications to use antipsychotics at times in persons who have a dementia disorder. So if the denominator is just persons with dementia, if that's the definition of the denominator and saying that those persons should not be getting

antipsychotics are 0 as the performance benchmark, it's – that's just – I don't – I don't think that's what any of the evidence guideline say or what practitioners think of practice.

Karen Johnson: This is Karen from NQF and we will certainly get to that point under scientific acceptability but the measure is constructed, so that certain patients, those with the diagnosis of schizophrenia, bipolar, Huntington's, or Tourette syndrome would be excluded from this. So there are – the measure is constructed with some – some exclusions, if you will.

But we are, you know, 15 minutes of not much time to talk about these measures but does anyone else on that committee have any questions or concerns about the evidence underlying the measure? OK, Gwen, do you want to go ahead and talk about –

Gwendolen Buhr: The comments about the numerator statement. I think that it should be more broad in terms of what kinds of conditions people may have if they are on the antipsychotic, that would help I think.

Karen Johnson: OK, that would go to reliability in terms of the specifications. So, was there any other concerns about the specifications in the measure?

Helen Kelman: Well things – lot of times, we see psychotic symptoms but really can't diagnose a clear schizophrenic condition or bipolar, so we give them a diagnosis of something like psychiatric disorder NOS and they have – they do respond to antipsychotic medication and very much need it for the – that's one example.

Karen Johnson: And was that Julie speaking just now?

(Helen Kelman): Oh no, this is (Helen Kelman), I'm sorry.

Karen Johnson: All right, are you with PQA?

(Helen Kelman): No, I – I was part of the PCPI workgroup and I'm with the APA also.

Karen Johnson: OK, just so, we don't know everybody's voices yet. So if you wouldn't mind, this is the first time you've spoken, we would appreciate if you just let us know who you are just so we can get to know people's voices but and I would – we didn't tell you that beforehand.

We also want to make sure that the committee, you guys, my committee members are a little bit quiet today. So, let's make sure the committee members get in the discussion points they need to and they will go to developers if they want a clarification. So Gwen, do you want to go back, is there anything on reliability that you wanted to bring up or – or point out from the preliminary eval.

Gwendolen Buhr: So from the preliminary evals, the two people who filled out the survey thought there was moderate reliability and moderate validity. So, I think that they did submit some testing of the measure with some different health plans and I'm just not – I'm not sure how we know – we know that they found the people with, you know, whatever they did.

I mean they found people who had dementia. They found people who had the dementia medications and they combined those together and they found that both together, resulted in more patients than either one by itself. My question is, how do you know there aren't more people with dementia that don't have a diagnosis and don't have medications. One of the other measure was giving evidence that people who are incorrectly diagnosed with dementia and people who aren't diagnosed with dementia, it should had been.

So, I think that there are likely missing a lot of people who have dementia and are on medications and don't have a diagnosis but I don't know. The data they submitted can't uncover that.

Jordan Eisenstock: Gwen, May I add one thing there. This is Jordan.

Gwendolen Buhr: Yes.

Jordan Eisenstock: Sorry to interrupt but just because it's related to your denominator statement there which I agree with. I was also a little bit concern about the – and this is more to note for revisions than may be for the live meeting but the denominator statement of the end or a prescription, you know, as we know medications aren't always just given for FDA indications but sometimes based on evidence and I think that we might not have a fair presentation of denominator for patients around these medications for some other indication like traumatic brain injury and I just thought that might be worth noting to.

Gwendolen Buhr: That's true. I had – I had patients who come to me with no diagnosis of dementia, no evidence of dementia but they're taking dementia medications. And some doctors are prescribing the medications to prevent dementia even though there's no evidence for that. So, I think you're going to grab patients who don't have dementia using their specification.

Karen Johnson: Is this something more of the developers would like to address very quickly?

Julie Kuhle: Hi, this is Julie Kuhle with PQA and we understand that some of these medications might be used for other diagnosis or purposes other than dementia but we think that – that's, you know, a relatively small group. Let's remember that these – these measure would be used for health plan. So, it's a prescription drug plan that really only have some diagnosis data and prescription claims data.

So, we know that dementia is underdiagnosed using ICD-9, so by using the medication and requiring a certain quantity of medications, so there's a chronic abuse, we thought that we would – were capturing more of the patients with dementia. So you know – you know it's kind of that risk and reward conflict and we felt that by including both the diagnosis and the medications that are pretty specific to dementia.

Although, you know, we understand that they are use for other things too occasionally. We thought that was the best. You know as far as the comments earlier about the medication or this measure is not ever reaching 0. So, there will be patients that need to

take these medications, the antipsychotic medications that don't have the exception diagnoses and do have dementia and its appropriate use. And we understand that, so – and we do understand that this measure should never reach 0.

But when we did our testing, we found a large variation, you know, 14 to 20 percent of the population that were taking these medications, the antipsychotics, without any of the diagnoses that were stated and that according to our denominator, had dementia. So, we felt that was a pretty high rate, so we felt that there is little for improvement here. So, I hope I addressed some of those questions.

We understand it's not a perfect – it's never a perfect measure when you're just using medication claims looking at a large body of population. I will say that it's – this measure was never intended for a facility level because we understand that – especially some nursing home facilities have a mix of a, you know – you know whether it's an Alzheimer unit or something. So, looking at a comparative population between facilities would not be the intention of this measure.

Gwendolen Buhr: But doesn't it say later that, you're – in the future you're going to stratify by a facility – at the facility level or something like that?

Julie Kuhle: I don't think so but if you see that, please let me know.

Gwendolen Buhr: OK.

Karen Johnson: This is Karen from NQF. I had a couple of quick questions in terms of the testing. You talked about the – we expecting to see both ambulatory and nursing facility patient, did your testing include both types of patients or did you only look at folks in nursing facilities?

Julie Kuhle: No, it was – it was nursing facility and outpatients. So prescription claims data and this is part D data, so that would be people that are in skilled nursing facility and then people who are ambulatory as well.

Karen Johnson: OK great and just to remind people on the meeting, in case you forgotten over the – the lovely summer, when we asked for reliability and validity testing, we – some developers that they could provide that testing at either the data element level or the measure score level and you guys for your reliability testing had presented some information about data elements. We need to see that sort of thing for what we call the critical data elements.

Those are the data elements that go into the measure and so I know you have addressed the dementia diagnosis and of course the – the pharmacy data but what I didn't see address and I'm wondering if – if there's any information out there, it's how good are the diagnoses for those folks that are being excluded from the measures, the schizophrenia, the Tourette, bipolar?

Julie Kuhle: You mean – so we excluded those four things based on a CMS measure at the time. So, we try to align our measure with the CMS measure. If your question is the ICD-9 codes that aligned with those?

Karen Johnson: Right, right are – are those good – are – are those pretty good? Are you, you know properly excluding those people like you want to or?

Julie Kuhle: Yes, I guess I'm not quite sure what your question is, those ICD-9 codes aligned with the diagnosis that we want to accept.

Karen Johnson: Right so.

Gwendolen Buhr: So –

Karen Johnson: Go ahead Gwen.

Gwendolen Buhr: OK, I was thinking I could maybe clarify. So, you show on page 15 the number of people with dementia from one of the plans and from the second plan. And you know – you show the number of people with the medication marker for dementia and so what I guess we would like to see is the number of people with the exclusion diagnoses before –

Julie Kuhle: Oh OK.

A committee mGwendolen Buhr: The provision diagnoses and then …

Julie Kuhle: We do have that information so.

Gwendolen Buhr: OK.

Julie Kuhle: OK, thank you.

Gwendolen Buhr: And one question I want to just propose and I don't know I think it might mess it up because of the logic of the outcome which relates to people with dementia but if you take all people who are taking antipsychotics for whatever the exclusions you want.

                  I was thinking that you would then get a larger number of people. Probably, a lot of them have dementia, but it wouldn't be – it would no longer be a dementia measure. It would just be on antipsychotic measure. But that to me seems a bit more valid, and until you try to link it with antipsychotics are bad for people with dementia. What do people think about that?

(Christy Taglin): Hi, this is (Christy Taglin). There is absolutely is a measure that does exactly that already as a CMS quality measure.

Gwendolen Buhr: Oh, OK.

(Christy Taglin): Yes. So this is the difference about these particular measures that it is focused on dementia and antipsychotic use.

Gwendolen Buhr: Yes.

Female: OK. In the interest of time, are there any other questions about the scientific acceptability criterion from the work group members? Any more concerns?

And then, Gwen, if you want to just summarize very quickly the feelings about usability and feasibility.

Gwendolen Buhr: OK. Usability and one person said, moderate, and one person said insufficient evidence. And feasibility, both said, moderate. So, I mean, the stuff comes from the pharmacy claims data, whatever, that you know, easy to get, I guess. And so that would be feasibility.

(Selena Ahwahnee): I think I was the one – this is (Selena Ahwahnee). I think I was the one that put down the insufficient data for the usability. And largely, my concern was about the taking – it was basically on do we really know why people on certain medications and back to the statement of the people using things off label for different reasons. And you're making a lot of assumptions based on it.

And so, it really gets back to I believe it was the developer who stated that it was going to a small percentage that they thought was custom common to this area. And I'm just not sure how small that actually is.

Karen Johnson: OK. And this is Karen. And just so you know (Selena) that, that is a very important point and that is probably what we would classify under validity. So, I think you got the right thing, just in the wrong slot essentially. So –.

(Selena Ahwahnee): OK.

Karen Johnson: And I guess, maybe another point on usability. What we're looking for on usability is used in public reporting and in internal QI programs. Since this is a new measure, we don't necessarily expect the measure to be in use, but the developer I believe did provide a rationale of why they think it would be usable.

(Selena Ahwahnee): This is (Selena) again. I'm just a little confused because I thought that even though that would be under validity that would impact whether or not meaningful.

Karen Johnson: You know, I guess you're right. Sometimes, it's almost like it would fall under both. So, I take that back –

(Selena Ahwahnee): I think – I think –

Karen Johnson: -- I think it would be both, especially if it could be misunderstood. If the results could be misunderstood by the public.

Female: Right. And I think that there's a lot of potential for the results to be misunderstood, for the reasons that we were talking about in the beginning that, I think that already because of the CMS initiative to reduce antipsychotic, that people are talking about not using antipsychotics at all in people with dementia.

And even though we all agree that, that isn't – that isn't the right thing, that's sort of the impression or the whatever. So you know what this measure was out there that would strengthen that notion and then as not all the people are in the numerator and the denominator, it's going to be even more misunderstood.

Female: OK, great.

(Julie Ann):   This is (Julie Ann).

Female: Oh, go ahead (Julie Ann).

(Julie Ann):   Yes. I would agree and this – I would agree with the last comment. And this is only coming as the wife of a geriatric psychiatrist who is giving what the stay in and stay out. I'm really worried about the unintended consequences that this – as it relates to selection of facilities or those kinds of things.

So, it'll be great if we could exclude, if there was something more in exclusion criteria where we could better capture the patients we really – we really want to make sure are not (standing) psychotics for lack of a better term.

Female: And – so this is the measure developer and this questioning of what additional diagnosis those might be. Because the evidence really is pretty strong that patient's with dementia on antipsychotics don't do better. They have poor outcome.

So if we're trying to get at the larger group with schizophrenia and bipolar disease, you know, what other exclusions would be included knowing that there's always going to be variability in patients that are going to need these medications, even if they are excluded using certain diagnosis.

(Julie Ann):   Unfortunately, not being clinical, I can't answer that question. I'm wondering person – physicians on the call that even if that's realistic – can provide input as to whether or not that's realistic.

Female: So, and this is the measure developer again, (Julie). And you know, one of our concerns is that we're trying to not take such a broad breast stroke of saying antipsychotic use is bad. And we're not just looking skilled nursing facilities, we're talking about the whole population of the ambulatory patients that are getting medications. And we know that there's a high rate of prescribing.

So, you know, what we really intended with this measure was to define a smaller group of patients where we know that the evidence shows that there are poor outcomes for new

patients are on antipsychotics. So, instead of just saying, let's decrease antipsychotic use, we really were intending to narrow our focus to – I think a population that maybe is more fragile and where the measure might be used more importantly.

(Julie Ann): That's helpful. Thanks.

Karen Johnson: OK. This is Karen. I hate to cut it off this really interesting discussion, but because of time, I think we do need to go on to our next measure. And actually, the next two measures are put forth by AMDA and they are very similar. So, I'll have (Julie Ann) go ahead and start us off on 2091, if you would.

(Julie Ann): Sure, 2091 is – it's just the indicators of dementia without a diagnosis, long stay. The numerator is – and it's really the percentage of nursing home residence over 65 or older with persistent indicators of dementia and no diagnosis of dementia.

Numerator, number of adult patient 65 or older who are included in the denominator that they have persistent signs and symptoms of dementia, and who do not have a diagnosis of dementia on any MDS assessment within the last 12 months.

Denominator is the total of all long stay residents in the nursing facility who have had at least two MDS assessments which may be in admission – may be in admission, annual correlates, significant change or significant correction assessment during this quarter, and who do not meet the exclusion criteria.

Exclusions, resident on hospice, comatose with delirium psychotic disorders, hallucinations, anxiety, manic depressive, PPSP (inaudible) schizophrenia, this is a facility level measure and it is a (inaudible) measure.

In terms of importance to measure and report, there was a lot of information provided indicating that dementia is under diagnosed. And three of the work group members helped impact was high, one, moderate. And three of the work group members felt that performance gap was high and one, felt moderate.

Of four, agreed that – that three that disagreed important to measure report. Some comments around mixing dimension, cognitive impairment and suggesting that cognitive impairment should be used, and that the title should be changed. And another comment around perhaps the statistical data on Alzheimer's disease wasn't – is current as it could be.

Karen Johnson: OK. Any discussion points around impact? OK. You want to go ahead to performance gap?

(Julie Ann): Oh, I think I mixed impact and performance gap.

Karen Johnson: Oh, OK.

(Julie Ann):   There was – there was a lot of information given suggesting that dementia's under diagnosed. Three work group members felt like the performance gap was high, one, moderate.

Karen Johnson: OK. Any discussion around gap? And generally, on the call, what we'll do is just have a very brief talk about these criteria and if nobody has concerns, we'll just get off of them quickly, and spend our time on the things that are more confusing or where you have more concerns. So that said, let's go on to evidence.

(Julie Ann):   So evidence –

(Selena Ahwahnee):   Actually, can I – I just – really quickly. This is (Selena Ahwahnee).

So I'm the one that brought up the issue of the cognitive impairment versus dementia, and how well is dementia actually diagnosed in some of these facilities. And also, some of the – some of the data if I remember correctly that was presented was really on cognitive impairment rather than dementia. Does anyone else have concerns – I mean, that was really one of my biggest concerns is – are people being labeled given the appropriate label?

Eric Tangalos: This is Eric Tangalos with AMDA.

I think, taking a step back to use cognitive impairment would be just an absolute mistake here. The BIMS identifies the dementia states as mild, moderate and severe. And the biggest problem we have is not making the diagnosis. This measure was all about at least getting to a dementia diagnosis and letting people slip away with cognitive impairment. This is really not what this measure is intended about, both for short and long-term stay.

(Selena Ahwahnee):   That's fine – I mean, just in my opinion, I agree with you. I'd rather people have a diagnosis of dementia than have an appropriate assessment. I'm just not entirely clear who's making the diagnosis and is that accurately reflects what is being guiding us.

Jackie Vance: This is Jackie Vance. And perhaps I can answer that, the – and I'm with AMDA, that the way the measures are assigned, and I'm sure we'll get into that or questions with that. When you have these two consecutive BIMS scoring and you have not had a diagnosis of dementia, only the practitioner leaves decision can make that actual diagnosis. Then we want to see it actually show up on the BIMS, which means the nurses now transcribe that diagnosis.

A nurse may not make that diagnosis. Nurses don't diagnose. So then we know that the physician has come in and done a medically necessary visit and did that differential diagnosis.

(Selena Ahwahnee):   OK. Yes, I just wasn't sure exactly how – what transpired and as who was responsible for that. Thank you.

Jackie Vance: You're welcome.

Karen Johnson: (Selena), this is Karen. When you were talking about some of the studies, talked about MCI, were you referring to the things that were noted in the evidence section or in the earlier sections of the form? (Inaudible).

(Selena Ahwahnee): I don't remember, but I can go back to it and look.

Karen Johnson: Maybe at some point. I was just curious.

(Selena Ahwahnee): Yes, I know. It was mainly because it was within – it may have been even within the importance of the measure that impact –

Karen Johnson: OK.

(Selena Ahwahnee): -- performance. And so, it mentioned it and then, I don't even think they said, mild cognitive impairment, just cognitive impairment. And both of those terms were being used, it seemed as if not necessarily interchangeably, but how much that information specifically applied to this, I wasn't sure.

And I, you know, if you're going to bring up that – if you're going to bring up cognitive impairment, then that potentially should be reflected in what we're – I wasn't sure exactly what we were going after, whether or not it was dementia, cognitive impairment or kind of a mix of both.

Karen Johnson: And do you feel better about it now, after the developer has given you some (inaudible)?

(Selena Ahwahnee): Yes, I do. I do. Thank you.

Karen Johnson: OK. OK.

(Christy Taglin): Yes, this is (Christy Taglin). Just to clarify that, just a tad bit more on the cognitive impairment measure that we are using now is the new – in the new MDS 3. It's a validated measure of cognitive status, but I think Eric said that we're using the severe cognitive impairment. So that's very clearly identified in the measure definition, in the numerator and denominator.

The exception to that is if the resident is so cognitively impaired that they cannot complete the BIMS interview, the – in which case, we are going – we have harmonized with the other CMS measures that use the staff assessment of cognitive impairment.

So, cognitive impairment is very much a part of getting to this underlying mis – under diagnosis of dementia, but we're very you know, specifically, setting those cut off points to articulate those most severely cognitively impaired patients that we're looking for. So, maybe that clarifies it a little bit better.

(Selena Ahwahnee): Yes, it definitely does.

(Christy Taglin): OK.

Karen Johnson: Do you want to go ahead and is there anything else to talk about in terms of the evidence presented?

(Julie Ann): (One-C).

Karen Johnson: The evidence?

(Julie Ann): So, under evidence, for the – based upon decision or – four people who rated this gave it a yes for quantity. Two, rated high. Two, rated moderate. For quality, one, rated high. Three, rated moderate. Inconsistencies, one, rated high. And three, rated moderate. One comment was there were no studies directly in the long-term care study.

Karen Johnson: OK. Any discussion around those points?

Eric Tangalos: Well, this is Eric Tangalos again. And we've already hinted at the cognitive impairment scale which is on MDS 2.0 and which is still imputed for those that are very, very severe. And when we looked at the CMS data that's been published on how much dementia is actually in the nursing home. It's from the MDS sources that used the MDS 2.0 computational score for the cognitive impairment scale. So we know that dementia is quite rampant in the nursing home.

Karen Johnson: This is Karen from NQF. I did have just a couple of questions real quickly. In terms of quantity, I noticed that you know, 242 was the number that was mentioned in the submission form. And I'm just curious as to where that number came from? And then, I just couldn't tell myself exactly if your evidence basis, a lot of randomized control trials, or well designed observational studies, or so, can you just give us a little bit more information about actually, the quantity and quality and consistency of the body of evidence?

Jackie Vance: Sure. This is Jackie. What we did was we used some very specific Peer Review general articles and then, within those articles, looking at the references, then pulled up the study to expose. Going by the instructions and the tutorials, where it says, don't (know if) the articles, the actual studies.

So then we pulled up the studies within those articles and looked at the evidence was there. And only the ones that had to do with what we were looking at – I mean, we weren't looking at the pharmaceutical studies or whatever within those articles themselves. And so, that's what – that's what took us. I think the longest was trying to pull all of that up.

Karen Johnson: OK. And so you're saying, you had to do that piece. There wasn't already systematic reviews out there that you could pull from.

Jackie Vance: Right. Because, yes, there wasn't or there wasn't the time to go in it and do that within the time to submit this as well. So and like you said, we had to go with what evidence exist

that you're not going to find randomized control trials in the long-term care setting. It's just not ethical.

So we have to apply whatever data there is to for our elders or to the subject of dementia and what we can. We can find numbers, pulling them up from MDS and things like that, but – or persons on certain types of medications by pulling pharma data from like large pharmacy providers for a long-term care.

But often, in this environment, you have to piece me all your evidence, because nobody's supporting trials in nursing homes.

Karen Johnson: OK. And then, just one other question. I just want to make sure I understand piece. You also use a guideline recommendation and this guide – this recommendation has not been graded. Am I correct?

Jackie Vance: The AMDA guideline is set up to be graded. It is being graded at the moment actually. This is in the national guideline clearinghouse.

Karen Johnson: OK.

Jackie Vance: And we have a grading system that was developed and approved by our board. And the guideline is in the process of being graded – will be graded prior to – if we move to the next step, if we'll be having the live meeting.

Karen Johnson: That will be really interesting to hear at the in-person meeting.

Jackie Vance: Yes.

Karen Johnson: Thank you.

Jackie Vance: Yes.

Karen Johnson: Committee members, do you have any other questions for the developer about the evidence or any other discussion points amongst yourself?

OK, let's go on to the reliability and validity testing, the votes were presented.

(Julie Ann): So, a scientific availability of measure properties, you had reviewer who said yes, three, who said no. Under reliability, one, who rated reliability as moderate, two, as insufficient. Validity, three, insufficient.

Karen Johnson: OK.

(Julie Ann): And we had comment about excluding psychosis because people with dementia commonly have psychosis.

Karen Johnson: OK. The members who said no on or insufficient evidence for reliability or validity, would you like to maybe expound a little bit on why you chose that rating?

(Selena Ahwahnee): (Ahwahnee) again. I was one of the ones that put the insufficient, cause I wasn't sure who was doing it, and how reliable and valid that would then be. So I feel much more comfortable now that the previous person told me that this was going to have signoff from a physician.

Karen Johnson: OK, great. Any other concerns about it, either reliability or validity? And just so everybody is clear on this, NQF says that if developers can provide evidence, data element validity, we do not ask for data element reliability information, so that these developers have done.

                   I think, one question – this is similar to some extent to the question that we had about the previous measure in terms of the critical data element and you – you spent quite a bit of time talking about the validity of the MDS itself. And also, that some of the different scales had been to the can and that sort of thing.

                   Where I got a little confused and I'd appreciate just a little bit of clarification is – you talk about at the beginning of the section, MDS validation done, comparing MDS data to a gold standard. Then later on, you talk about a phase validity process with the Delphi process. So, what is two different ways of assessing validity or can you just clarify that for me?

Jackie Vance: I'm sure, (Christy) could answer this even better than I, but according to Dr. Saliba who I spoke with, who puts brand and one of the developers of the MDS 3.0, they say there were two different testing methods. So there was the gold standard and then the other one. So, they worked with two different testing methods.

Karen Johnson: OK, great.

Jackie Vance: Do you have anything to add to that (Dr. Tai Kwan)?

(Dr. Tai Kwan):   No, that's correct, Jackie. The MDS items went to an extensive validation process that included both phase validity, content validity, and item validity. So, it was a very extensive process.

Eric Tangalos: Yes, and this is Eric again. Just this month, that whole process is published in the Journal of the American Medical Directors Association.

Karen Johnson: Oh, great. Is that something that you've made available, the citation to us?

Eric Tangalos: Why again, that's a – that's a new citation. I know it's out there because I wrote the editorial for all four parts.

Karen Johnson: OK.

Female: Yes, it wasn't – it wasn't coming out until this month, so I could not include it with this. I included the last articles, you'll see them cited. You'll see Saliba as the first author in those.

Karen Johnson: Yes.

Female: We certainly can add the references if that's something that you need or you know, I can still (inaudible) with me.

Eric Tangalos: Yes, and then, BIMS went through its own process and had that published first. And then, was incorporated in the MDS and had its incorporation published again.

Karen Johnson: Right. And I believe I saw the BIMS citation on there. So –

Eric Tangalos: Jordache would've been the lead author.

Karen Johnson: OK. (Julie Ann), do you want to finish this up, on usability, feasibility, any concerns with this item?

(Julie Ann):   Sure. On usability, one work group member rated high, three, moderate. And feasibility, one, high, two, moderate, and one, insufficient. No, no comment. Well, there was one comment about having a measure on cognitive impairment. Are there any concerns about the usability or feasibility as measure?

Karen Johnson: OK. Sounds like not. And again, I'm rushing you just because of the clock here. (Selena), you have the second measure which my understanding is it's processed the way – the same as this one, only, for short – stay patients.

(Selena Ahwahnee):   It's pretty much the same, so –

Karen Johnson: OK.

(Selena Ahwahnee):   -- I cannot just breathe through the top part.

Eric Tangalos: And this is Eric, I'm going to have to sign off. Thank you very much.

(Selena Ahwahnee):   So yes, there is some work –

Karen Johnson: Yes.

(Selena Ahwahnee):   So why don't we just go ahead and go down to the importance of the measure. There was a significant impact, with three being high, one being medium. In terms of addressing a performance gap, two, high, two, medium. And there was a comment that there was – the data really wasn't divided into short stay versus long stay.

                   But I think everyone still came to the same conclusion that those are – those are significant impact. In terms of the evidence, there was agreement that there was

significant evidence. Those in terms of quantity, quality and the consistency with pretty much agreement of high and medium. In terms of science acceptability, there's a little bit of variability on that in terms of the reliability and validity.

And I think I was the one that was in determinant. It was for the exact same reason. So I feel much more comfortable if I was the one that said the, in determinant. I'd be – hello –

Karen Johnson: Yes.

(Selena Ahwahnee): -- I'd be happy to move those over into moderate…

Karen Johnson: Yes.

(Selena Ahwahnee): -- in terms of usability, everyone thought it was usable, highly is or usable except for – I believe that was me as well. And it's for the same reasons, so I'd be happy to move that over into moderate in feasibility that was agreement as well.

Karen Johnson: OK, great. So again, I'm going fast. I think we've covered most of the concerns that you guys at least have brought up. I guess my one question and this is just curiosity on my part. The developers have splitted up into short stay versus long stay patients. Does that ring appropriate to you as the committee members?

Female: Can the developer tell us why?

Female: Absolutely. It's because how the assessments are done in long-term care based on your stay status and your payer status. So, we actually had to divide that up and that harmonizes with the other CMs measures and how they are done based on short stay and long stay. Again, based on how the MBS assessment is done in time.

Female: OK. So it – OK.

(Christy Taglin): Yes. It's very much – it's (Christy) – very much how CMS has designed all of the other new quality measures. They are separating out the long stay population from the short stay population, because they're you know very different entities. And so, we followed that – that same logic, virtually using the 100 days perspective payment assessments versus the OBRA assessments. So, it's in – it's in the definition.

But if you're not familiar with the MDS language, you're going to miss that clear definition between short stay and long stay.

(Selena Ahwahnee): So, one question if I could just follow that up. I mean, it really seems like this is more of a reimbursement question. But in terms of quality, is there a significant difference between what's provided in those two? Between short stay and long stay?

Female: I'm not so sure I understand your question. I apologize.

(Selena Ahwahnee):   So most of the information if I remember correctly, it was kind of all sort of lumped together and wasn't divided out that way.

Female: Well, one – the patient population may be slightly different. But also, the timing of the practitioner visits are different based on medical necessity.

(Selena Ahwahnee):   Right.

Female: So you'll probably have less long stay patients who may go undiagnosed versus long stay patients, unless it's pointed out. For example, a short stay patient is going to have the – that BIMS assessment done at day five and day 14, and then again, at day 30 versus the long stay patients whose going to have it done at day 14, day 30, and day 60.

So, it has a lot to do with how they're going to be – when they're going to be assessed and how they're going to be seen. And we want to be sure, we cover them.

(Selena Ahwahnee):   Oh, OK, yes. OK, I see.

Karen Johnson: And can you tell me, I think lost the train there. You said that you would expect to possibly see it less (inaudible).

Jackie Vance: I'm wondering, I know, I'm just – I'm wondering – I would love to – I want to see – I want to see what's going on out. So, one of the things I'd love to see just out of curiosity, they were not measuring it. I would love to see it if you're having more physician practitioner visits for the short stay patients because of the acuity status, I'm wondering if you're going to miss less versus the ones that are coming in and only seeing every 60 days.

I'm curious. I'm just curious. And as now things we're kind of measuring, and I wonder, maybe, that we can look at that secondarily. That somebody, a geek like me would love to see that. You know, I'm just wondering if we would see that.

And by separating that out into the short stay and long stay, if ever, anytime you want to do a secondary study or information to pull from.

Gwendolen Buhr: So, this is Gwen Buhr and I'm just answering your question about if it makes sense to separate them. And it makes total sense to me because the short stay and the long stay patient is somebody totally different. The short stay is somebody you know rehabbing for something. And they might go home or, or they may transition into a long stay person. And they may have more delirium because they've just been in the hospital or you know, whatever the long stay person is a much more stable patient.

And I'm thinking, maybe it's opposite of what you're saying Jackie and that the short stay person – that physicians are just concentrating on their knee rehab and they're going to miss more dementia. And then the long stay person, they're more interested in the whole patient.

Jackie Vance: That could be too.

(Christy Taglin): I was going to say that as well. I agree. It's (Christy). I think that the – you know, we're very concerned about this short stay population. It's important to break them up because they might not be focusing on that. They – you're right, they are focused on fixing that hip fracture and you know, this is a perfect opportunity to get that diagnosis so that when the resident goes home, they get the right care. So they don't end up back in the nursing home for the long stay. So I think that variable might be the case.

Jackie Vance: I would just love to see the difference.

(Christy Taglin): Yes.

Karen Johnson: OK, great. If there's no other discussion points that are just burning, I think we probably do need to move on to our next measure. So, any other burning issues on these two measures? OK, great.

Now, we're going to switch gears a little bit and go to the PCPI measures on dementia and are late discussant on this is Tina Cronin. Tina, are you on the line?

OK. I guess Tina got called away and didn't make our call. Would anybody else who evaluated the measure want to give a shot at doing the lead discussant role on this one? Don't all volunteer at once.

OK, why don't I take a shot at it and I'll just lead folks through. This measure is title neuropsychiatric symptoms assessment. And the measure looks at the percentage of patient with a diagnosis of dementia for whom an assessment of neuropsychiatric symptoms is performed and results reviewed at least once in a 12-month period.

It is a process measure specified for the physician level. And data sources are claimed other electronic clinical data, EHRs, and registry data. So in terms of impact, looks like the reviewers were fairly high on impact. Most thought that there was a high impact because there are a large – a large number of patients with dementia and because behavioral problems are common in folks with dementia. So, impact looks like – that was well demonstrated.

In terms of performance gap, all four evaluators evaluated this measure as either high or moderate. And I think I'll stop there and see if there are any comments from our committee on either impact or performance gap. Any questions or concerns?

OK. In terms of evidence, I think one of the questions that I had and we're going to get back a little bit to Section IC.1 where we ask the developers to tell us what the structure process outcome relationship is. So I think, if I understand this measure correctly – well, you know what, instead of me telling you what I think you said, maybe the developer can just very quickly tell us how you expect, you know, what is the outcome that you expect if this symptom assessment is done?

Diedra Joseph: Hi, Karen. This is Diedra.

Karen Johnson: Hi.

Diedra Joseph: So, the neuropsychiatric symptom assessment measure just for clarification is actually paired with the management of neuropsychiatric symptoms measure. But with regards to your question about structure process, outcome relationship – the measure is a process with which leads to the accurate identification of significant or dangerous behaviors and triggers, appropriate prioritization of interventions, and development and targeted support, and educational strategies for caregivers.

Karen Johnson: OK. Thank you.

Diedra Joseph: You're welcome.

Karen Johnson: So, the evidence base is I think you based it on a clinical practice guideline. Actually, looks like three of them, one from APA, one from a California work group, and then, also a Canadian guideline. So, looks like the committee for the most part thought that the evidence was there for this measure. Although, one commented that we need more information regarding the type of studies and the limitation of the studies.

So, I guess sometimes, when guidelines are used as the base, it may not be exactly clear, exactly the kind of studies that were done and how many were done, and that sort of thing. So I noticed that grades were assigned for the APA guideline and for the California work group. And you tell us what the grades mean.

So I guess, one of the first things from the committee that might be worth asking is, are these well known and trusted guidelines? Are you happy with these guidelines?

Jerry Jones: Let me comment. This is Jerry Jones and I'm – I was on the development team on the – on the work group as a coach chairperson.

Yes, these are particularly the APA guideline is one of the highly recognized and effective guidelines. And maybe (inaudible) is and the California work also. So yes, these are recognized respected guidelines.

Karen Johnson: OK. And can I direct that question to the committee members. Is everybody happy with these guidelines in terms of feeling – feeling that these are (inaudible) the commonly used reputable guideline?

OK. You guys are really quiet.

Female: I agree that definitely the APA guideline is a common guideline that you would look to. I have not personally heard about the other guidelines, so.

Karen Johnson: OK. So the APA guideline, it looks like the – if I understand the grading correctly, it's based on category one. It seems to be a category one recommendation, which I believe

means recommended with substantial clinical confidence. Is that correct? Am I understanding this correctly?

That would be a question I guess for Diedra.

Diedra Joseph: I'm sorry, I'm just trying to go through the mission form to find that information for you.

Karen Johnson: OK.

Diedra Joseph: Oh, yes, APA guideline category one is recommended with substantial clinical confidence.

Karen Johnson: OK. And can you explain, I mean, does that help us really understand what the body of evidence is underneath, or is that more just a confidence about the recommendation, or can you just throw some light on that for us?

Diedra Joseph: Yes. The grade that you're looking at is actually the grade for the recommendation. The body of evidence was not graded in the APA guideline.

Karen Johnson: OK.

Female: Now, one on the other guideline, the grade is level three which means, opinions of respected authorities based on clinical experience, descriptive studies, or reports of expert committees. What she said wasn't a good enough evidence.

Karen Johnson: Yes, NQF guidance does want measures that have a firm evidence base. And generally, opinions don't – don't come to that level.

Somewhere, I read and I'm not sure if it's in this particular submission or if it was in another one, and I'll just read you the quote. APA guidelines will select moderate clinical certainty that usually requires Class II evidence or a strong consensus of Class III evidence. Does that sound right to you in terms of how APA guidelines are constructed and how they're used? And I'm throwing that out to the developers or to the committee.

Jerry Jones: Yes. I think that is consistent with their process.

Karen Johnson: OK.

Jerry Jones: This is Jerry Jones.

Karen Johnson: OK. And I think that the question then that I would be curious about is since it would usually require a Class II evidence or Class III evidence, or a strong consensus of Class III. Could you provide or find for us a description of what those two classes mean?

Jerry Jones: We can do that. I mean, if some of this evidence for the assessment part is really speaking to the incidence and prevalence on neuropsychiatric symptoms and persons with

dementia, as well as the association of those symptoms with any kinds of outcomes. And I mean there are some studies in that area.

So, we can – we can pull that between –

Karen Johnson: OK.

Jerry Jones: -- now and the next meeting.

Karen Johnson: OK. What we'll probably end up doing is opening up your submission forms that you can add that to the submission form.

Jerry Jones: Sure.

Karen Johnson: OK. From the committee, any questions or concerns about the evidence underlying the assessment measure?

OK. Reliability or validity, again, these are a brand new measures. They have not yet been tested and – but we do have information about the specifications of the measure under Section 2A1. So I guess any questions from the committee or any concerns about the specifications for the measure.

And it looks like, in terms of the comments, both people who – actually, all three people who looked at this measure gave a rating of either low or insufficient evidence or for reliability and validity. And one comment was, is there a definition for dementia that is used? And then, a question about a couple of the ICD-9 codes. So, would anybody like to expand on that and tell us what you were thinking about that?

Jerry Jones: Yes. This is Jerry again.

The definition of dementia was based upon a long list of codes that include all forms of dementia. So that there – we were trying to be broad in our definition here. And the five or six common causes of dementia which are on the list – I don't have a list in front of me now, but there's a long list that includes some rarer causes of dementia also.

So we think that list of ICD-9 codes and 10 codes will include the vast majority of persons who have dementia. The purpose of these measures is to the same in persons who do carry a diagnosis of dementia, are they having a new and psychiatric symptom assessment done.

So the denominators is persons who have – who have some diagnosis of dementia. This will miss persons who don't get the diagnosis at all, but that was not the purpose of this – of this measure.

Karen Johnson: OK. Committee members, do you have any questions or concerns particularly after that explanation from the developer?

Gwendolen Buhr: This is Gwen Buhr.

I put insufficient after the guidance from you all and maybe I got confused. I thought that I was supposed to do that if it was an untested measure, and they didn't put the testing results. I was supposed to put insufficient.

Karen Johnson: OK, and that's fine.

Gwendolen Buhr: OK, yes.

Sam Fazio: And this is Sam. I did the same thing.

Karen Johnson: OK. Any – it looks like there was a little bit of concern about those usability and feasibility. A couple of people thought feasibility was low. Usability, 1 percent thought it was low. Is there any comment that you want to air about that?

Let see and under feasibility, I'm seeing a comment. I'm not sure why they response to 4a1 to 2 is generated during provision of care. Are they not arguing that it frequently is not? What that item is getting at is that something that is generally done in – during provision of care.

In other words, the collection of the date – what NQF is asking with this question is to collect the data for this kind of – for this measure does the provider have to do something totally different or is this what is should be normally done. So, you're right that they are arguing – they are arguing that isn't, but if it were, it would be done during the normal provision of care.

Male: Absolutely. These are symptoms that persons with dementia commonly have but tend to be overlooked because practitioners have not recognize the importance of them and so they may not record them, assess them and into that (inaudible).

Karen Johnson: So, this is going to require chart abstraction, is that right?

Male: I'm trying to see how (inaudible).

Karen Johnson: This is specified for claims or other electronic data. So, they have not specified it for paper records. Although, arguably if it's in EHRs, there might be some manipulation required to get the information out of EHR.

Male: Some of this clearly would be described, for example, their long list of activity disturbances, their mood disturbances like depression for example, which is going to come up separately. They're starting perceptual disturbances. So, there are and in the record means of recording these set of neuropsychiatric census, some of them more easily than others.

Karen Johnson: OK. Thank you for that. I think as we go through again a reminder that often it is more difficult to find strong evidence on measures that are more distal to the desired outcomes and generally an assessment would be considered quite distal.

So, committee members, I would like to do this kind of think about that as you think more about this measure and then all the other measures in the project, just keep that proximal, distal question in mind.

Male: Can I make on other comment about this – I think maybe even this more germane. One approach to this as recommended in a numerator is the use and instrument – a valid instrument like the (NPI) or dementia signs and symptoms scale. So, we have – so the practitioner can use an instrument and certainly that would be more ease – could be found more easily or could just recorded in the medical record. Either approach would be appropriate.

And in nursing homes –

Karen Johnson: OK.

Male: -- could use the MDS.

Karen Johnson: Jordan, are you ready to tell us a little bit about the paired measure – the measure 211 that they are pairing with the assessment measure.

Jordan Eisenstock: Sure. Ready to go.

Karen Johnson: Thank you.

Jordan Eisenstock: Oh, sure. So, this is 2011 Dementia Management of Neuropsychiatric Symptoms also from the AMA-PCPI. The submission is a new one. Its description is percentage of patients regardless of age with the diagnosis of dementia who have one or more neuropsychiatric symptoms were received or were recommended to receive an intervention for neuropsychiatric symptoms within 12-month period.

So, this means that the numerator is patients who received or were recommended to receive an intervention for neuropsychiatric symptoms within a 12-month period and the denominator is outpatients regardless of age with the diagnosis of dementia who have one or more neuropsychiatric symptoms. There were no exclusions. This is at the clinician level of analysis and is a process measure, and as you mentioned, it's paired with 2010 neuropsychiatric symptoms assessment and management.

Move on to impact, I assume at this point, looking through the preliminary survey, it looks like – it's about everybody who responded which were four of the members where in general agreement, three rated the impact high, one moderate. I didn't have the writing – that's not me, but I agree with it, you know, there were several numbers reporting large patients with dementia and probably one of the pertinent numbers is also cited there that 33.7 percent of the patients were recommended for behavioral problems.

As far as another – just important thing that was written in the impact part, the fact that he prevalence of neuropsychiatric symptoms based on at least one community sample was 40 to 88 percent which is obviously a wide rage but this show that this has potential for high impacts.

I don't know if you want me to stop there to see if there's any comments or not.

Karen Johnson: Yes, please. Any concerns with impact? OK, let's go on to gap.

Jordan Eisenstock: With regard to the performance gap, the number is or as far as the preliminary survey were exactly the same three for high and one for moderate. A lot have the same information there. This is actually where, you know, the fact that only 33.7 percent of the patients were recommended shows that there is some room for improvement and probably that was – that was probably the most important piece of data that was – that was in the measure summary. But again, I think there's a lot of agreement there. So, I'm not sure I should go too much further.

Karen Johnson: OK. If no burning question, let's talk – I think for this one probably – let's concentrate a little bit on the evidence and how strong we prove that evidence is and then any other things that may come up from the committee members.

Jordan Eisenstock: Sure. So, moving on to evidence, this is a process and as far as the breakdown with quantity, quality and consistency, there was one outlier in consistency that suggested insufficient evidence but everyone else was still in the high or moderate range. With regard to quantity, it was two and two for high and moderate, quality four in the moderate range and consistency, one high and two moderate.

As far as the pertinent structure-process-outcome relationship, I think that this was pretty well stated in the measure, and under 1c.1 neuropsychiatric symptoms of dementia have associated with accelerated cognitive decline, increased functional impairment, decreased mean survival time, increased comorbid conditions, increased danger to self, increase danger to others, increased health service utilization, high risk for institutionalization and greater caregiver stress and burden which helps make the conclusion that management of neuropsychiatric symptoms, therefore, is critical to providing high quality care to demented patients, and I think that this was well supported with several references in that section.

As far as quality, I – let me just see what I had written in some notes. There was just the information provided based on the Third Canadian Consensus Conference, as well as – what's the other one – oh the California Workgroup for Alzheimer's Disease Management.

Again, I think everyone was in relative agreement that this was well documented and supported information. Let me just see here, there was one comment that I wanted to remember to address which actually was – it was written under usability. Someone had written the comment would be more useful to have pharmacologic and pharmacological tabulated separately.

I assumed that was a misprint and maybe based and the only reason I mentioned it in this section is because the grade assigned to the body of evidence was level 1 for non-pharmacologic treatments and level 3 for all others. So, there was a separation there.

Should I open it up for comments?

Karen Johnson: Yes, please.

Committee members, any concerns about evidence? Again, just because of time, this one a little different I think than the other and that I didn't see the APA – an APA guideline for this measure. The Canadian guideline looks like the grade may be based on opinion for one piece of it but then grade the fair evidence level 1 at least one RCT for another one of the guidelines.

OK, Jordan, let's see if we can take maybe just a couple more minutes and then finish up this measure if we can.

Jordan Eisenstock: OK, sure. I –

David Hackney: This is David Hackney. I'm afraid I have to sign off.

Karen Johnson: OK. Thank you, Dr. Hackney.

Jordan Eisenstock: So, reliability and solidity received insufficient from everybody who reviewed it. We've gone over those reasons before. As far as usability, one high, two moderate and one insufficient, some of the comments there which might be worthwhile just in the short period of time left –.

So, I assumed the first comment was with regard to the fact that the Neuro PI module is going to be launching and is going to be part of the larger quality improvement curriculum and then again, there is – there's not about pharmacological and I assumed non-pharmacological tabulating separately which it might be the reason for the insufficient, but I don't know whose comment that was specifically.

As far as feasibility, again, three moderate, one insufficient here and it looks like for similar reasons that maybe that person rated it insufficient, and I think it might be worthwhile just up in there open for comments.

Karen Johnson: So, I'm not hearing much in terms of concerns with this measure from our workgroup members. Is that correct?

Jordan Eisenstock: Speaking for myself, I would say, no I'm not that concerned. I think it was – I think it's OK.

Karen Johnson: OK, great. All right. Our last measure is measure 2016, Screening for Depressive Symptoms, and Sam this one is yours.

Sam Fazio: (No problem). So, this measure was brought forward by the American Medical Association, the Physician Consortium for Performance Improvement with the description of the percentage of patients regardless of age with the diagnosis of dementia were screened for depressive symptoms within a 12-month period. The denominator statement would be all

patients regardless of age with a diagnosis of dementia, no exclusions and it is a process measure.

So, just a little bit about the impact statement, the measure has intended to encourage the detection of depression given its high prevalence in dementia patients and subsequent impact. The detection of depression is essential for early intervention and proper management. As far as the committee members, it looks like there were four committee members who evaluated this measure. There of them said high impact and one for the moderate.

Karen Johnson: Great. So, go ahead and go to performance gap.

Sam Fazio: OK. So, the performance gap here as far as the ratings, we have two folks with high impact, two folks with moderate impact, (with Karen). Some of the information that they presented to us in the performance gap is quoting a recent study that found 63.4 percent of patients were given formal instruments for depression screening, and another study that talked about 73 percent. So, it looks like folks had higher moderate impact for the performance gap.

Karen Johnson: OK, committee members, any concern with performance gap?

OK, how about evidence?

Sam Fazio: OK, so, go on to evidence here, the process measure still I'm talking about the quantity, quality and consistency here, so it looks like sort of across the board, everybody voted moderate on quantity, quality and consistency. Someone commented it could have more studies if we look this information that was provided related to this study.

It looks like they were related to AAN practice guidelines, APA practice guideline and California workgroup although there is not specific studies that were discussed each of these guidelines that, you know, a significant number of studies that were cited in the reference session of those guidelines. It looks like APA had a 554, AAN 147 and California workgroup over 400 articles.

Karen Johnson: OK. Any concerns or questions about the evidence presented? And I would just ask the committee members to think about the ratings that we asked you to do for quality and I'm – it is more difficult with guideline evidence. So, it looks like – I don't think the APA guideline was graded.

Look it was – I'm seeing the grade and then also the California workgroup guideline, the evidence was not graded on that. So, I guess this is back to the questions of the developer if you could find for us the definition of class 2 and class 3 evidence which is what the APA guidelines are – with what according to the statement on page seven, that would be very helpful to understand what class 2 and class 3 evidence means, what that is.

So, I – and makes – let me just make sure, I – you understand what I'm asking for just a statement that say, you know, class 3 evidence reflects, you know, randomized control trials or you know, made an analysis or whatever class 2 means and whatever class 3 means.

Male: Yes, we can do that.

Karen Johnson: OK. That would be very helpful. OK, how about reliability and validity?

Sam Fazio: Well let's see here, so the numerator statement was patient who was screened with depressive symptoms within 12-month period and some information was provided about clinical qualitative approaches that could be used. I think that from our ratings here, we have three that are insufficient and same thing for validity with a comment there is a lack of data.

Karen Johnson: OK, so again, this measure had not been tested. I guess is there any questions about the specification of the measure? Is that information we do have? Any questions or concerns about this specification?

Diedra Joseph: Karen this is Diedra. If I may, I just wanted to clarify that dementia measures were submitted at the request of NQF staff for time-limited endorsement and that's why there's not testing. All right, data just – so that the workgroup members are aware and the testing project for these measures are set to start in October of this year.

Karen Johnson: Oh, great! That's good to know. Yes, what we've asked the developers to do is if their measures are recommended for endorsement, that would a 12-month endorsement which we give them time to do the actual testing. So, and they have agree to that. It sounds like you're well in your way to getting the assessment.

OK, Sam, how about usability and feasibility?

Sam Fazio: Sure. For usability, the measures currently used as part of the dementia measures group in the CMS decision quality reporting system program and the PCPI believes that the use of these measures in quality improvement is a beneficial way to get the scientific data which will improve physician performance, committee members – we have one high and two moderate and one person who selected insufficient.

And when we look at feasibility, it looks like we have two moderate and two insufficient with a note that says, "Need more information on the sub-criterion." I wonder if more information on questions related to feasibility need to be completed, seems like it might have been a little slim.

Karen Johnson: Did the committee members particularly the ones who rated usability or feasibility is having insufficient information, would you care to expand on that or maybe give the developers any idea about what you would like to be seen here?

OK, I think with usability, just a reminder, this is a new measure. So, we wouldn't necessarily expect to see actual use although it is part of 2012 PQRS, and it looks like it's going to be part of an AAN module in terms for QI use. The developers did provide rational and plan for use.

Jerry Johnson: May I ask for clarification? This is Jerry. On the usability, are you asking about the feasibility of its future use or experience with its past use?

Karen Johnson: Is it – is a measure that – has already been endorsed for a while. Generally, with would expect to see some information about how it's actually being used out there in the world, and we don't necessarily require that for our or expect that really for new measures. However, since it is part of the 2012 PQRS it actually is in use. So, again, a lot of new measures wouldn't be able to say that.

So, it already is in use and – for public reporting and then, again, you've talked about the plan for use for QI purposes. So, I think you've been very responsive to the usability question.

Jerry Johnson: OK.

Karen Johnson: Yes, I don't know if I answered your question or not.

Jerry Johnson: Yes, you do.

Karen Johnson: OK.

Jerry Johnson: Yes, I think so.

Karen Johnson: On the feasibility questions, the one that a lot of times I think is a very useful piece is 4c, the susceptibility to inaccuracies, errors or unintended consequences, and oftentimes that is more difficult to answer if it's a brand new measure but as you, you know, you may not know what some of the potential errors or unintended consequences are but, you know, if this measure is endorsed and then it comes around three years down the road for maintenance particularly on 4c we would love to see any analysis or information that a developer might have just in terms of potential problems with the measure or things that maybe didn't quite workout the way they thought it would and maybe have that resolved.

So, that's part of the – what we're trying to get to the feasibility and, again, it's a little bit less – it's part of the answer if it's a new measure at this one.

Jerry Johnson: Thank you.

Karen Johnson: Steering committee, any other questions, concerns, comments about this measure?

OK, if not, we're not going to turn the call over to Suzanne who's going to finish up for us.

Suzanne Theberge: Thanks, Karen. (Amy), can you open the lines and we'll start our public comment session.

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

Again, that was star one for question.

There are no phone questions at this time.

Suzanne Theberge: OK, thank you. All right. Next steps, we would like to ask the committee members to continue their review and start looking at the other measures, the ones that were not in this workgroup, we'd like you to review all those measures in time of the in-person meeting on October 3rd and 4th, and we're looking forward to seeing you in just a couple of weeks.

It's hard to believe that so soon, but that will be here in DC. You should have received some information about that from our Meetings Department with information about making travel arrangements and your hotel arrangement. If you do not receive that e-mail, please let me know, and we'll fix that for you.

And for the developers as well, you folks should register also whether you're planning to attend by phone or in-person. We do want you to register so we make sure we have facilities setup to accommodate you.

So, with that said, the committee should go ahead and continue reviewing measures. For the developers, we'll be in touch with you regarding any updates that you need to make for the measure forms after this call, and we'll look forward to seeing you all in a few weeks at the in-person meeting in our office.

Male: Thank you. Bye-bye.

Suzanne Theberge: All right. Thanks everybody.

Male: Thank you.

Male: Thanks.

END