NATIONAL QUALITY FORUM
+ + + + +
NEUROLOGY PHASE II
STEERING COMMITTEE
+ + + + +
THURSDAY
OCTOBER 4, 2012

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 9:00 a.m., David Knowlton and David Tirschwell, Co-Chairs, presiding.

PRESENT:

DAVID KNOWLTON, MA, New Jersey Health Care Quality Institute
DAVID TIRSCHWELL, MD, MSc, University of Washington Department of Neurology
A.M. BARRETT, MD, Kessler Foundation

WILLIAM BARSAN, MD, University of Michigan Health System
JOCELYN BAUTISTA, MD, Cleveland Clinic
RAMON BAUTISTA, MD, MBA, University of Florida, Jacksonville
GWENDOLEN BUHR, MD, American Medical Directors Association

GAIL COONEY, MD, FAAHPM, Hospice of Palm Beach County
JOHN DUDA, MD, Veterans Health Administration
JORDAN EISENSTOCK, MD, CPE, UMass Memorial Health Care
SAM FAZIO, PhD, Alzheimer's Association
RISHA GIDWANI, DrPH, Stanford University

Medical Center
DAVID HACKNEY, MD, Beth Israel Deaconess Medical Center

MICHAEL KAPLITT, MD, PhD, Weill Cornell Medical College
DANIEL LABOVITZ, MD, MS, Montefiore Medical Center
THERESE RICHMOND, PhD, CRNP, FAAN, University of Pennsylvania School of Nursing
JACK SCARIANO, JR., MD, FAAN, private practitioner
PETER SCHMIDT, PhD, National Parkinson Foundation
RAJ SHETH, MD, Nemours Foundation
JOLYNN SUKO, MPH, Virginia Mason Medical Center
JANE SULLIVAN, PT, DHS, MS, Northwestern University Feinberg School of Medicine
FREDRIK TOLIN, MD, MBA, FACS, Humana, Inc.
MARY VAN DE KAMP, CCC-SLP, RehabCare
SALINA WADDY, MD, National Institutes of Health

NQF STAFF:
HELEN BURSTIN, MD, MPH
KAREN JOHNSON, MS
SUZANNE THEBERGE, MPH
JESSICA WEBER

ALSO PRESENT:

SUSANNAH BERNHEIM, Yale-New Haven Health System CORE*
KERI CHRISTENSEN, American Medical Association
AMARIS CRAWFORD, American Medical Association
ELIZABETH DRYE, Yale-New Haven Health System CORE
KATE GOODRICH, CMS*
ANU GUPTA, AMA-PCPI
LEIN HAN, CMS

KENDRA HANLEY, AMA-PCPI
JERRY JOHNSON, University of Pennsylvania*

DIEDRA JOSEPH, American Medical Association*

HARLAN KRUMHOLZ, Yale*

ROBERT PLOVNICK, American Psychiatric Association

SAMANTHA TIERNEY, American Medical Association

*Participating by teleconference

C-O-N-T-E-N-T-S

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

(8:31 a.m.)

CO-CHAIR TIRSCHWELL: Good morning everybody, if we could settle down a little bit, I think we are going to go ahead and get started. Thanks for showing up on day two.

Karen is going to start us off with a little recap of day one.

MS. JOHNSON: Thank you, David.

So, hopefully, you guys all remember, we had a very interesting and exciting day yesterday. We looked at 12 measures, and of those 12, three were recommended by you for endorsement.

So today we are going to look at some additional measures. Most of them today will be dementia measures, and one measure that has already been endorsed by NQF, the stenosis measurement in carotid imaging studies.

So all of these measures are put forward by AMA-PCPI. So to start us off

today, we are going to ask folks from AMA-PCPI to give us maybe a little five-minute introduction overview to your measure set.

MS. TIERNEY: Good morning, everyone. Thank you for your time and the opportunity to offer a few brief remarks about our measures, the 10 measures that are presented for you today.

The measures come from two different sets of ours, one in our radiology measure set and one in our -- and many, nine measures, in our dementia measure set.

I do want to just point out for you a few highlights of the PCPI measure development process, just so you have a sense of what goes into the development.

All of our measures are developed through multi-disciplinary, cross-specialty work groups. We place a strong emphasis on developing measures that are based on clinical practice guidelines.

We subject all measures to a 30-

day public comment period, and then we review all the comments we receive with the Development Work Group for further consideration by the Development Work Group, and make measure modifications, where necessary.

We also subject all of our measures to the membership of the Physician Consortium for Performance Improvement for vote and approval. This is a very important step in our process, given that the membership of the PCPI is very large and diverse. It includes over 170 medical specialty societies, state medical societies, and health care professional organizations.

I will just speak a few minutes about -- or one or two minutes about the dementia measures in particular, and then I will turn it over to my colleague who can give you a slight overview of the one carotid measure that is presented for you today.

The dementia measures are the

result of a collaboration among the PCPI, the American Academy of Neurology, the American Geriatric Society, the American Medical Directors Association, and the American Psychiatric Association. They are the results of a year--plus long collaboration to develop measures to improve care for patients with dementia.

As I know that you have noted from several of your calls and probably in looking at the submission forms more closely, there is a challenge with the evidence base for the measures. We do at the PCPI strongly believe in basing our measures on clinical practice guidelines, with a reliance on trying to develop measures that are based on principles with the strongest recommendations and based on the highest level of evidence.

Unfortunately, in the area of dementia, there is no a strong research base, and so we identified measures that would be important to improving the quality of care for

patients with dementia, but we recognize that there are some challenges with the evidence base. But we strongly believe that the measures do have a great potential for benefit, and strongly outweigh any harms and, I think, in some situations, maybe many, given the weak evidence base, we would ask for the possible exception to the evidence requirement for NQF's criteria.

I will also point out to you that the measures are up for time-limited endorsement, because they have not yet been tested for reliability and validity, but they meet all of the other criteria that are required for consideration by NQF under those criteria, and we are in the process of planning a testing project, and I think we will begin later this month.

So that is short and sweet, hopefully, and I will just turn it over to Diedra on the phone and see if she has any extra comments to add about the stenosis

measure, in particular.

MS. JOSEPH: Good morning, everyone. This is Diedra Joseph from the AMA-PCPI. Thank you for the opportunity to introduce the measure.

The stenosis measure, number 0507, was developed by a Radiology Work Group. The Work Group was developed in conjunction with the ACR and the NCQA.

The measure was developed by this group and approved by the PCPI membership in 2007, and received time-limited endorsement status from NQF in 2008, and the measure is supported by two clinical practice guidelines and was tested for reliability and validity, along with the three other AMA-PCPI radiology measures, which also originally had PLE status and were recently reviewed and granted full endorsement.

So that is our brief introduction of those measures, and we welcome any questions you have throughout the Steering

Committee discussion. Thank you.

CO-CHAIR TIRSCHWELL: Thank you very much. With no further ado, I think we will go ahead and start on the first measure.

The scheduled first measure is 0507. David Hackney is going to talk about stenosis measurement in carotid imaging studies.

DR. HACKNEY: Okay. This is, as you heard, an AMA-PCPI proposal for stenosis measurement in carotid imaging studies. It establishes the useful goal of encouraging standardized reporting methodology for patients with carotid stenosis using the NASCET approach.

It has been improved and previously endorsed, and as you will hear, I have some problems with it, but the nature of the problems and the potential value of it are such that I would suggest we renew the endorsement, but ask the developer to make some revisions that would better match the

apparent goal and avoid some of the problems I see now.

I am going to start with the note that this standard uses stenosis severity as the only required on a report of carotid imaging in a patient -- well, in a patient with carotid imaging, and it ignores things like ulceration and plaque composition which, I think, most radiologists would consider important parts of the report and, depending on what is going on with the patient, may be more important than the stenosis.

So it kind of implies that, if you have reported stenosis, you have done everything you need to do in characterizing the severity of the vessel abnormality. So I will make that note briefly. Obviously, it would be a big production to add other elements to the report and test them, validate them, and bring them forward, but it is something I would encourage them to think about.

I am most worried about the denominator. It seems to assume that the severity of carotid stenosis will be relevant for everybody who undergoes imaging of their cervical arteries.

You typically look at the carotids but not the vertebral arteries, for example, when you do an ultrasound, but -- and it is technically possible to do that with MRA, but the way MRA is usually done and the way CTA is done, you get all the cervical arteries.

So if you do a study, even with no interest in the carotid arteries, according to this, you have to report the severity of the carotid artery stenosis. I think that is what this means. That is what we have been doing, in any case, because we think that is what it means.

Now that is sort of a meaningless distraction for the people taking care of the patient when the issue isn't carotid disease, to begin with, but there is also this issue

about perhaps referring a lot of patients who are asymptomatic for carotid disease but have carotid stenosis getting endarterectomies or stenting, because they have asymptomatic carotid stenosis.

This is a big issue right now. There is a big debate about whether there is any need to any intervention in an asymptomatic patient, if they can undergo medical therapy, but -- and as you do older patients, most of them have some carotid artery disease. So measuring it may lead them into a therapeutic pipeline, where they don't belong.

So there is a potential harm to it. There's lots of other reasons people get neck vessel imaging. As I said, it could be trauma, looking for dissection, pseudoaneurysms, tears in the vessels, hematomas, neck AVMs, tumors, and not all of those people is the carotid artery of any interest unless you saw a totally unexpected -

- it is significant, because it is there in carotid stenosis, usually isn't significant, just because it is there, again unless the patient is symptomatic.

Now applying the NQF standards, there really isn't evidence that doing NASCET stenosis reporting will have a positive effect on patient care. There is good evidence that, in symptomatic patients, the stenosis as measured by the NASCET technique does predict stroke risk, but I don't know of any studies, and certainly the developer didn't indicate any, that show evidence that including that in the report has an influence on patient care, and it is a technical issue that is going to come up when we have to go point by point through the criteria, but I think the link between stenosis severity and stroke risk in patients who have symptoms is strong enough that it is plausible to think that documenting that formally would be a useful thing to do.

So with that --

CO-CHAIR TIRSCHWELL: Right. We should be focusing on the evidence first. I guess, David, just as a question to the evidence, is there evidence that, if they don't use the NASCET method, that the reported stenoses are inaccurate?

DR. HACKNEY: There are different reporting -- There are different methods that have been used, and you get different numbers if you would use a different method. So first of all, if you -- There is another big one that uses what the diameter of the carotid bulb you think would have been in the absence of the stenosis. That has got obvious problems about deciding how tight that would be, but the important thing is, if you use that, you don't get the same number as you do if you use the NASCET.

So if you report it without defining the method you are using, people don't know how to use it, and most of the data that has been developed for stroke risk

prediction from carotid stenosis severity is with NASCET.

So if you don't use NASCET, then it is hard to know how to plug whatever number you get into the existing data. So to the extent that you care about how tight the stenosis is, doing it by NASCET gives you by far the biggest database on which to base your subsequent clinical decision.

So that part of it makes perfect sense. Proving that, having that in the report, makes the patient better off, is the sort of thing we were discussing a lot yesterday, that that is tough to do.

So as I said, I am willing to accept that it makes so much sense that you need that number, that I am not worried about the fact that you can't prove that putting it in the report matters, because it would be a weird study for someone to do. But it is going to be a point when we talk about where there is evidence and where there isn't.

CO-CHAIR TIRSCHWELL: I have one more question on this that maybe you can answer, and it sort of goes to evidence about this way of reporting carotid stenosis, which is the ultrasound question where, you know, as far as I can tell -- and I looked at the consensus report about ultrasound reporting.

Ultrasound is kind of a different animal, and they report it in these ranges of stenoses. That is what the recommendation was from the consensus statement that they reference, and this range of stenoses is based on peak systolic velocities, I believe, and it has nothing to do with the distal carotid diameter.

So it sort of feels like they are forcing that one in there, too, and it doesn't quite fit with the title of the measure. I don't know if --

DR. HACKNEY: Yes, that is true. You really can't -- In the vast majority of people, you can't see the segment of the

carotid that is your base for calculating the stenosis severity with ultrasound. It is out of the window. Some people you can, but many, many people you can't.

So you can't do it that way. But I think doing a standardized method would also be useful for ultrasound, but you can't really do what NASCET did, and there are papers that try to link ultrasound measurements to NASCET method measurements so that you can derive -- you can use that database of information about stenosis severity, but it is an extra step that you have to make in order to get there.

CO-CHAIR TIRSCHWELL: Well, but the peak systolic velocities that they do recommend using do correlate pretty well with these ranges of NASCET stenosis. So I guess it just feels a little not quite consistent to be including the ultrasound thing in there, when you are not really doing exactly what the measure is called. I don't know.

DR. HACKNEY: Yes, I think the

alternative on that issue would be to have a separate measure just on ultrasound, I guess, but I think there is a logic in grouping these two together. Again, if you ask for the direct evidence, again, that the report alters therapy as opposed to the information alters therapy, that -- and there isn't going to be, I don't think.

CO-CHAIR TIRSCHWELL: Ramon, Daniel, then Michael.

DR. R. BAUTISTA: It sounds like this particular measure has to do with standardization of a certain process. I guess the question is, is there any need to standardize this. Will this result in better patient care and, really, what is the evidence that standardization of this kind of a process is actually good?

DR. HACKNEY: There is very good evidence, as I was saying, that using standardized criteria for assessing the severity -- there is a great deal of data on

predicting stroke risk using that; and because there is more than one way of measuring carotid stenosis, that gives you different numerical values.

If you want to use the largest database with the most studies and the most evidence to stratify stroke risk based on stenosis severity, then NASCET is the method that you want to use. So there is very good evidence that you can predict stroke risk using NASCET, and that if you use a different method, you get different numbers.

So you can't use the NASCET database, but can you prove that having that in the report changes therapy? That, I don't think there is any data on.

DR. R. BAUTISTA: Then a follow-up, though: Is this process or protocol valid across different procedures discussed, like ultrasound, MRA, CT angiogram, etcetera?

DR. HACKNEY: So you do the same thing with CTA as you do with MRA, and those

have been validated between the two of them. As we were just discussing, you can't do this in most people with ultrasound. So you use a different criteria entirely, but people have related the ultrasound criteria to the CTA and MRA criteria.

MS. JOSEPH: Excuse me. This is Diedra at the AMA-PCPI. May I make a comment?

CO-CHAIR TIRSCHWELL: Why don't we let the committee make their comments, and then you can respond to them.

MS. JOSEPH: Okay, thank you.

CO-CHAIR TIRSCHWELL: Daniel?

DR. LABOVITZ: I am stroke neurologist, and I read a lot of ultrasound reports, and I appreciate the notion of standardization. But ultrasound remains probably the most commonly done test to assess for asymptomatic or symptomatic carotid stenosis.

When I look at this measure, I am reminded of -- I think it was Sesame Street

where they ask you which one of these things is not like the other.

CO-CHAIR TIRSCHWELL: Hey, they are not getting funding anymore, by the way.

DR. LABOVITZ: Yes.

CO-CHAIR TIRSCHWELL: Just want to bring that out. Big Bird is out.

DR. LABOVITZ: I think I am little concerned here that the measure is a measure -- - It is sort of a standard of convenience: Let's put all of these things together, because they are measuring carotids. But they use very different means.

MRA is different from CTA, is different from angiogram, and all the NASCET data comes from angiogram. That is what was used to establish the standard. That is what measuring the proximal and distal portion of the internal carotid artery is from.

Carotid ultrasound, when assessing stenosis, looks at flow velocity. It is also useful for looking at plaque morphology, which

angiogram can't do. This measure doesn't have anything to say about that, but it does ask carotid ultrasound to do something which it does very poorly.

Maybe this is a question for the rest of the committee, but certainly, I think, might be a question for the developers. Do you think that including carotid ultrasound in this standard is useful and valuable? Is there perhaps an unexpected downside to this, forcing the ultrasonographers to generate a report which isn't valid, maybe even misleading? Do you have to have it?

CO-CHAIR TIRSCHWELL: Michael?

DR. KAPLITT: Yeah. I have two questions to the points you made earlier. One is on this point of reporting the degree of stenosis.

Putting the ultrasound question aside, which I agree that there is serious sort of structural concerns about how related ultrasound is and whether the data supports

it, if every study that has been done, whether it is asymptomatic treatment of symptomatic treatment, shows that the benefits really occur above a certain level of stenosis. Right, then I guess do you really need a study to show that, actually, reporting in a consistent way the level of stenosis is actually beneficial to outcome, if you have already shown in many, many studies that the degree of stenosis, not just severe versus moderate versus whatever, but the actual percent stenosis, affects the outcome. Right? So that is one question.

The second question is with regard to your concern, which I generally share, but with regard to concern about all studies. I guess the issue is: Let's say somebody is getting an MRA or a CTA because they have had a head injury and you are worried about a dissection because of the nature of their injury. It turns out they don't have a dissection, but they have a 90 percent carotid

stenosis. Isn't that what radiologists are supposed to do?

So I share your concern. Your concern is for the patients where they suddenly report like a 50 or 60 percent thing on a suboptimal study, and what do you do with that. But that is clinical judgment.

Same thing when we get an MR and you see some abnormality, some lesion that you didn't expect. Is it a tumor, is it not, whatever? So you got to work it up further. But I think that I am less concerned personally about requiring people to report it, as long as it is standardized.

I think, if it was not standardized and you required them to report it, that, to me, raises actually more concerns to some degree. So those are my questions.

CO-CHAIR TIRSCHWELL: Bill?

DR. BARSAN: Just along the same line as what Daniel said. For the developers, I just don't know how you can fulfill the

numerator statement if the only imaging study you did was a neck ultrasound, because the numerator says you are going to do measurements of distal internal carotid diameter, which we have just been told many times you can't do.

So I don't know. It just seems like a disconnect.

CO-CHAIR TIRSCHWELL: Yes, and I guess I will -- the Developer probably has a number of things to say, and I will suggest the possibility -- I think the title, actually, Stenosis Measurement in Carotid Imaging Studies, is fine, but perhaps the numerator has to be changed to something like what it is now, final carotid imaging study reports that include direct or indirect reference to measurements of the distal internal carotid artery as the denominator for stenosis measurement or, if the assessment was ultrasound, standardized criteria for reporting according to the radiology

guidelines, or something along those lines.

Does the developer have any comments related to all that?

MS. JOSEPH: Hi, this is Diedra at the AMA-PCPI. I just wanted to try to address your concern about the ultrasound.

Actually, in the numerator details we include a definition about the direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement.

I know that this was a point of discussion during the original review of this measure. So we actually were able to update this definition, hopefully, to address your concern.

The definition is that it includes direct angiographic stenosis calculation based on the distal lumen as the denominator for stenosis measurement, or an equivalent validated method reference to the above method; for example, for duplex ultrasound

studies, velocity parameters that correlate with anatomic measurements that use the distal internal carotid lumen as the denominator for stenosis measurement.

So I think that the reason why that definition was added was to address that concern. I think that perhaps Dr. David Seidenwurm, who I think is in the room there, could address your concerns more specifically.

CO-CHAIR TIRSCHWELL: So you are saying that you already included that in the numerator details, a different approach for the carotids?

MS. JOSEPH: That is correct. There is a definition in the numerator details that was --

CO-CHAIR TIRSCHWELL: I guess that is not reflected well in the numerator statement that is at the top of the page that everybody is paying the most attention to, and you might want to update that a little bit. Dr. Hackney?

DR. HACKNEY: Yes. It is 2(A)(1)(3), and it is part of the indirect language refers to, that the indirect is a way of saying in part that you are using something else other than actually measuring the distal carotid, but you are able to relate that severity to the severity measured using the CTA or MRA distal carotid method.

CO-CHAIR TIRSCHWELL: So, essentially, the ultrasound standards were used compared or set up compared to a NASCET approach and, thus, serve as a proxy for the ultrasound testing?

DR. HACKNEY: Yes, and there is good data on that, that you can derive the same -- you can derive equivalent numbers from ultrasound.

CO-CHAIR TIRSCHWELL: So that is great. It is already --

DR. HACKNEY: But you don't measure.

CO-CHAIR TIRSCHWELL: It is

already in there, and I guess we just need that reflected to some degree in the short numerator statement that is at the beginning of the measure, but I am sure the -- I am guessing the developer would be happy to make that change.

Any other questions or comments about the evidence from the committee? Jack, sorry.

DR. SCARIANO: If the actual standards are actually based on the MRA finding or the CTA finding, if that is the actual standard, then when actually someone has the ultrasound, what the surgeons are going to say is, well, you know, the actual standard, the actual standard that, actually, we can see, is on the MRA and also CTA.

I have this problem now, that actually, I don't know. If you have an abnormal ultrasound or do you need an MRA or do you need a CTA or do you need an angiogram? It is kind of up to the vascular surgeon.

So I think, in having this standard, it may actually confuse it even more.

CO-CHAIR TIRSCHWELL: You know, my personal perspective on that is that it is highly variable what surgeons require prior to doing endarterectomy, whether they are asymptomatic or symptomatic, how much they trust their local lab, how much they like their radiologist. So I don't think we can answer that or really even address it with this particular measure.

David, then Michael.

DR. HACKNEY: Just the asymptomatic versus symptomatic, I don't want to get lost in there, because I think that is a substantive issue. The technical thing about the ultrasound versus CTA, MRA, I think, they have dealt with, and they can update a little.

I think the significance of a given severity of stenosis is drastically

different in the asymptomatic than in the symptomatic patients. In the asymptomatic patients, at least at my place, what will happen to you if you have a carotic stenosis depends very much on whether you see a stroke neurologist or you see a vascular surgeon.

If you have a 60 percent asymptomatic stenosis and get sent to a vascular surgeon, you are going to get recommended to have that fixed. They are going to stent it or they are going to do an endarterectomy.

If you go to a stroke neurologist, they will say that, with proper medical management, your risk of stroke is so low that it is almost impossible to document that there is a method to make it lower.

So showing somebody who is asymptomatic as a 60 percent stenosis and putting him in a mechanical therapy pipeline, you have done that patient a disservice. That is the element that I am worried about.

I don't think there is any problem with reporting a standardized measure of stenosis in those patients, but I don't see that they are benefitting, and I think they could be harmed by it. That is my concern about the asymptomatic ones, and then it becomes tricky to know whether somebody is symptomatic or not.

CO-CHAIR TIRSCHWELL: I guess, seems to me that is not the radiologist's job at that point. Michael?

DR. KAPLITT: That was the point that I was going to make. Whether or not different groups of providers are intervening based on something -- I mean, this is a radiology reporting measure, and I think we are extrapolating many steps down the road.

I can tell you that a surgeon who already feels that ultrasound is enough for them to go ahead and operate on a patient is not going to be changed by this, and vice versa. A surgeon who doesn't feel ultrasound

is adequate and wants an anatomic study is not going to suddenly operate on people because we are standardizing the measure.

In fact, it could be the opposite, which is that you will get more consistent practice, because you are standardizing the measure, but I think we are extrapolating too much. I think ultimately the question from the evidence standpoint, which is the question we are still on, I think, is whether or not there is adequate evidence that standardizing the actual measurement has value.

I come back to the question I asked earlier, which is that, if the clinical data really does show that the degree of stenosis, not just the qualitative measure, actually influences therapeutic outcome, then isn't that evidence that there is value in actually standardizing it?

Now whether each modality in here is justifiable, I think, is a reasonable argument, but that is a separate question.

CO-CHAIR TIRSCHWELL: Okay. Any other comments before we vote on the evidence? Let's go ahead and vote then on the evidence.

MS. THEBERGE: Seventeen Yes; six No, evidence does not meet guidance; one No, insufficient information.

CO-CHAIR TIRSCHWELL: So check my math, but I think we continue to proceed here.

The next topic probably can be pretty brief: High impact. Yes. Okay, I propose we vote on that. Any other comments before we vote? Okay.

MS. THEBERGE: Twenty-one, High; three, Moderate.

CO-CHAIR TIRSCHWELL: And then is there evidence of a performance gap or an opportunity for improvement?

DR. HACKNEY: Again, yes, they present very good evidence that a lot of people don't do this.

CO-CHAIR TIRSCHWELL: Let's go ahead and open the voting. Go ahead and start

voting now.

MS. THEBERGE: Can everyone vote one more time? Okay.

Twenty-one, High; three, Moderate.

CO-CHAIR TIRSCHWELL: Moving on to scientific acceptability, starting with reliability. David?

DR. HACKNEY: This one, I was a little tough with. There are some precision problems in those reports, but I think there is good data that you can get precision that is good enough to make useful predictions.

So I think it is reliable, and the specifications are precise enough. I would like them to clarify some of the issues I was raising earlier about who is included in the numerator and denominator, but how you do the measurement is quite reliable.

CO-CHAIR TIRSCHWELL: Anybody have comments or questions on that? Let's go ahead and open the voting then. Go ahead and start voting now.

MS. THEBERGE: Nine, High; 15 Moderate.

CO-CHAIR TIRSCHWELL: Okay. then validity?

DR. HACKNEY: In their study where -- they reported that expert opinion was the criteria for validity, and the expert opinion strongly supported it. That is the only evidence of validity, but it was unanimous, I believe, among their experts.

CO-CHAIR TIRSCHWELL: Anybody have any questions or comments about the validity issues? Okay, let's go ahead and open the voting up. Go ahead and start voting.

MS. THEBERGE: Three, High; 20, Moderate; one, Low.

CO-CHAIR TIRSCHWELL: Moving on to number 3, usability.

DR. HACKNEY: I was at moderate. As I said, since it is an impact on clinical care by proxy, the measurement is useful to know. No data on the measurement being

included in the report is critical, but you could assume it.

It is understandable, with some confusion I brought up about the denominator; useful for public reporting, because it would have the effect of improving performance, information on performance, I would say, for a subset of the patients. But it could also be misleading for patients who undergo neck vessel imaging for other reasons, but meaningful, understandable, and useful for public reporting. I was at moderate for those reasons.

CO-CHAIR TIRSCHWELL: Any questions or comments from the committee? Let's go ahead and open the voting. Go ahead and start voting now.

MS. THEBERGE: Three, High; 20, Moderate; one, Low.

CO-CHAIR TIRSCHWELL: Then number 4, the last main criteria, feasibility.

DR. HACKNEY: Feasible

demonstrated by product of care processes. Yes, High.

CO-CHAIR TIRSCHWELL: Okay, let's go -- any comments or questions from the committee? Let's go ahead and vote.

MS. THEBERGE: Eighteen, High; five, Moderate; one, Low.

CO-CHAIR TIRSCHWELL: Then one last vote, which is on the overall suitability for endorsement. Any further comments or questions before we go ahead and vote? Okay, let's go ahead and open the voting.

MS. THEBERGE: Twenty-four, Yes.

CO-CHAIR TIRSCHWELL: Very good. Moving on to the next measure, Fred, first dementia measure, Neuropsychiatric symptom Assessment. Oh, I'm sorry. Jocelyn, go ahead first.

DR. J. BAUTISTA: Just a procedural question. So this measure was first endorsed in 2008 under time-limited endorsement, and it is still under time-

limited endorsement four years later, according to the --

CO-CHAIR TIRSCHWELL: I think they said that they had gotten full approval in the interim.

DR. J. BAUTISTA: It says time-limited status not yet removed.

MS. JOSEPH: I can address that. This is Diedra. We originally received time-limited endorsement status in 2008. At that time, with time-limited endorsement we were allowed two years to test the measure.

Additionally, we applied for an extension for EHR testing, and that was for one year, and the measure was submitted at the end of 2011 for review of the time-limited status endorsement with the testing data. However, because of the neurology endorsement maintenance coming up, we were asked to submit this measure for full review. So that is why the lag.

DR. BURSTIN: It is no longer

time-limited.

DR. J. BAUTISTA: So the measures that we approved yesterday for time-limited -- is that 12 months?

DR. BURSTIN: WE have changed that policy now. It is 12 months.

DR. J. BAUTISTA: But, potentially, they could extend it.

DR. BURSTIN: No. No more extensions.

CO-CHAIR TIRSCHWELL: All right. So back to dementia again, Neuropsychiatric Symptom Assessment.

DR. TOLIN: This is measure 2009, the first of the dementia assessments. This measure deals with the evaluation of neuropsychiatric symptoms in individuals who have dementia. Let me start over again.

Dementia assessment: This is a measure to evaluate the neuropsychiatric symptoms of individuals who are diagnosed with dementia, and it is meant to evaluate this

assessment being done at least annually or at least once a year.

In the numerator statement the assessment is divided into a couple -- three main groups, activities, moods, thoughts and perceptions, and there is a list generated in the numerator which is not meant to be exhaustive. It is just not an inclusive list, just more of an example list, listing a number of these, and there's also some suggestions, although not mandatory, about several types of scales that can be used which are commonly used in research settings.

In the denominator, it is all patients who carry a diagnosis of dementia, and this is not limited to any setting. So it can be either in a facility or living semi-independently. I would assume that dementia patients usually don't live completely independently.

As far as the question of outcome or how it is related to outcomes, this measure

is paired with measure 2011, which is the next measure we will be discussing, and it has to do with the treatment of the neuropsychiatric symptoms.

The evidence for this measure is not based on any trials, but is, in fact, based on expert opinion.

I will stop there, David.

CO-CHAIR TIRSCHWELL: Okay. Thank you.

DR. TOLIN: Oh, I meant to -- This is part of the PQRS. This is the dementia measures group of PQRS.

CO-CHAIR TIRSCHWELL: This is a physician level measure.

DR. TOLIN: Yes, physician. Sorry, I should have included that.

CO-CHAIR TIRSCHWELL: I will just add just a little bit about the evidence thing, and this is in the summary document. A couple of different recommendations statements were referenced as evidence, and

the process/outcome relationship included something along the lines of assessing neuropsychiatric symptoms leads to their identification, and then can trigger appropriate intervention.

So it is another one of those multi-step situations that the evidence is based one. I think, for this measure, as for many of the subsequent measures, specifics of evidence, types of trials, things like that, were not really included in the form that was submitted.

This came up again and again, I think, on multiple conference calls, and some of the recommendations were graded. Some of them weren't. Many were based, as you said exactly, Fred, on expert opinion.

Anybody want to add to that before we vote on the evidence? Does the developer have any comments before we proceed with a vote?

MS. TIERNEY: Yes. Hi, This is

Sam Tierney. I just would like to make a few comments about the evidence. The challenge has been in answering the quantity/quality question.

So as I said in the introduction, we based on measures on the practice guidelines, and many of those are evidence based. Actually, probably all of them we used are evidence based, and do some sort of review on the evidence, oftentimes also supplements it by expert opinion, where needed.

Unfortunately, the various guidelines that we have relied on for these measures from the American Psychiatric Association, from the Third Canadian Consensus Conference on Diagnosis and Treatment of Dementia, and from the California Work Group on Guidelines for Alzheimer's Disease Management -- they include some indication that they done a thorough review of the evidence, but that information is not available to the reader of the guidelines.

So where possible, we have tried to include any sort of information that might address to some extent the questions that were asked in the submission form, but I think ultimately the challenge is that, for dementia, there is not a very strong evidence base out there, particularly for assessment type measures or counseling measures, and those are unlikely to be subject to randomized controlled trials.

I know that NQF in their Evidence Task Force Report has recognized that some aspects of health care are more difficult to study with quantitative methods, particularly randomized, controlled trials, and that some process steps may be unlikely to be subjected to research.

so we believe that many of these measures may fall within that area. So I think, as you are voting on evidence, if you find that it isn't sufficient, which we recognize that, we might ask that you could

consider the exception to the evidence requirement.

So that is just all I will add. I don't know -- I think we might have Dr. Johnson on the line. I don't know if he has anything he would like to add about the evidence. Thank you.

CO-CHAIR TIRSCHWELL: thank you. Is there a doctor on the line that wants to comment?

DR. JOHNSON: Sure. This is Jerry Johnson. The evidence from observational studies --

CO-CHAIR TIRSCHWELL: Can you talk a little louder, please?

DR. JOHNSON: Yes. Yes, there is evidence from observational studies about performance gap, not evidence, that speaks to whether assessment itself leads to changes in outcomes. That evidence doesn't exist now, and I don't know that that type of a study will ever be done.

So I agree with the comments that were just made.

CO-CHAIR TIRSCHWELL: Okay. thank you very much. A.M.?

DR. BARRETT: I have a specific question to ask the developer about this kind of evidence. Dr. Johnson, this is A.M. Barrett.

With regard to these dementia measures, a question that came up about many of the aspects of quality clinical practice which are included in these measures, is that internal data may be available, since many of these measures may be overrepresented in quality settings, such as specialty clinics.

Therefore, there may be a possibility of assessing outcome data in comparing patients who have had these measures versus those who have not had them, to demonstrate the value of such assessment. Has that been performed?

DR. JOHNSON: No, that hasn't been

performed either, just whether or not -- What we know is that persons with dementia, who are documented to have dementia, seldom have a precise assessment necessary to make clinical decisions, in this case about neuropsychiatric symptoms. So we have that kind of data.

CO-CHAIR TIRSCHWELL: So that is the evidence of a performance gap.

DR. JOHNSON: Yes, and that is true across a variety of different settings, primary care settings as well as specialty clinics.

CO-CHAIR TIRSCHWELL: Okay, thank you. Michael.

DR. KAPLITT: So given that it is clear what the evidence vote is going to be, because the developer themselves, everyone of them, said there is no evidence on this point, but they have several times requested the exception -- so personally I would like to hear the argument for that. I would like to hear the argument for it, because we are not

going to vote unless someone raises it, because this is going to be, it sounds like, the same thing for like a lot of these coming after.

So I would personally like to hear what the argument is in favor of the exception, given the discussions of how the exception should be invoked yesterday.

CO-CHAIR TIRSCHWELL: So, Michael, you are asking the developer what their argument is?

DR. KAPLITT: Or even on the committee. Does anybody on the committee have an argument in favor of the exception that we should be discussing. I guess, if not, then that's it, but since the developer specifically asked for that, I would like to know what the argument is.

CO-CHAIR TIRSCHWELL: Let's get these few comments, and then I will specifically ask --

DR. KAPLITT: If they say there is

no evidence, so that's that.

CO-CHAIR TIRSCHWELL: Peter, Daniel, then Gwen.

DR. SCHMIDT: So if you accept that there is evidence that the differential diagnosis of different characteristics of dementia will inform -- it can be used as evidence to inform therapeutic decisions, then you have got what the UK NICE guidelines would refer to as a none or some criteria, which they classify as Class I evidence.

If no one is going to get that evidence based therapy based on there not being an assessment, but some people will based on their being an assessment, the UK would classify that as Class I. So you don't have to -- It doesn't necessarily mean there is not evidence if no one is going to get this therapy in the absence of this assessment, and some people will.

Many groups that assess evidence do not consider that to be a leap of faith.

So in the absence of doing this assessment, no one will -- If we believe that the evidence for differential therapy -- So for example, in Parkinson's Disease there is a poster at the Movement Disorder Society saying that people with higher executive -- people whose dementia or whose cognitive decline is more in the executive dysfunction domain have a higher incident of false.

So in Parkinson's Disease, if you assessed a higher level of executive dysfunction, those people would need more false counseling. You would do an OT home visit, things like that, and if they have a more generalized dementia, it is less of a risk.

So if you don't do that differential assessment, then you cannot -- you could give everybody the OT home visit, but there is a difference in the way that you would address the disease in these people based on this differential assessment; and if

you don't do the differential assessment, then you can't make that decision.

So no one would get the benefit of having the therapy tailored to the characteristics of their dementia, if you don't do the assessment, and some people will get it, if you do. So in the UK that would be considered Class I. Does that make sense? No?

CO-CHAIR TIRSCHWELL: I understand what you are saying. I am not sure that that evidence grading system is what we are working with here. Do you want to comment on that, NQF staff?

MS. JOHNSON: I do just want to remind you that, when we ask you to evaluate a measure based on evidence, we do ask you to look at the quantity, quality, and consistency of the body of evidence. So you have to be able to look at the submission and see what they have provided in terms of that.

CO-CHAIR TIRSCHWELL: Okay. Daniel?

DR. LABOVITZ: I enjoy beating dead horses. So just indulge me for a minute, because I don't think I will get much opportunity to say this again.

I think all of these measures have the same fundamental problem. This is an evidence based committee. I figured that out after about a minute, and I think everybody else did, too. It demands evidence and, if you don't have evidence, there are exceptions, but I don't think this is really a question for the developers. It is more a question for NQF.

There are going to be times when you come up with a measure where there is no evidence or no evidence that would meet the template that we use for evaluating these measures, and I would suggest that developers would have had a better time of it and could have made a better case if there were a process for that kind of measure, where you have no data but we have a compelling reason

to put this out anyway.

Instead, developers are forced to sort of twist through hoops, then finally today say, well, yeah, you are right, you know, there really isn't anything; and I hate to see them have to do that. I would rather see them put out something we can use.

CO-CHAIR TIRSCHWELL: Gwen?

DR. BUHR: And speaking about the exception, I don't think this one would qualify for an exception, personally, because it is not something like your analogy of the Parkinson's executive dysfunction. These neuropsychiatric symptoms are not subtle and, in my experience, the patient's caregivers are coming complaining of these things.

It is not something that -- and if they are not coming complaining, then fine, they are doing okay, I think, and we have already discussed they are not supposed to be on any psychotic.

So if everybody is assessing for

subtle things and there is nothing really to do about it except for complicated nonpharmacologic things, then why should we assess it? So I don't think it should be an exception.

CO-CHAIR TIRSCHWELL: Okay, round two of the comments. A.M.?

DR. BARRETT: First I would like to clarify that I am not requesting an exception, but I would disagree with what you said, respectfully. I think the rationale is what Michael asked for was a rationale for an exception rests on three arguments, the first being the value of assessment of adequate or clinically standard assessment and initial diagnosis. I think that many of us appreciate that and that, indeed, there will never be any randomized controlled study of that.

Secondly, the value of initial and repeated assessment of clinically standard symptoms such as neuropsychiatric symptoms in targeting care appropriately; and I wish I

could say that I had confidence that neuropsychiatric symptoms are always assessed adequately by history.

Unfortunately, having observed, for example, someone who came into a memory disorder clinic I was heading saying I have familial Alzheimer's disease, who after alone giving that history after five years of being treated in a specialty clinic, and actually most likely had depression causing persistent static cognitive symptoms. I don't believe that that is the case personally.

Again, I think there is an opportunity for the developers to present evidence of this based upon administrative data.

Thirdly, the argument is of public health -- Sorry, that second argument also is, I think, the argument that Peter was addressing in which other countries have had different standards for evidence.

Thirdly, the argument is for

public health. Of course, the burden of Alzheimer's Disease is going to be a significant problem for all of us, and we need to -- I was going to say instantiate -- We need to establish certain behaviors in our clinicians in order to take advantage of treatments that may reduce the burden of Alzheimer's Disease, and specifically, for example, cholinesterase inhibitors and other treatments have been suggested for neuropsychiatric symptoms as contrasted with other symptoms.

So that is the argument: Initial diagnosis; targeted treatment; public health. Again, I think it is an opportunity for presenting data in the second category.

CO-CHAIR TIRSCHWELL: Jordan, then Gail and peter.

DR. EISENSTOCK: I just wanted to add one thing really briefly. I am the lead discussant for the next one, which is 2011, and it is paired with this one. So I think

the head is about to get cut off of the body of the one that I am about to talk about and, because of that, I just felt it might be necessary to add one piece of information about the performance gap, which might lead to us invoking the exception.

Again, to A.M.'s point, I don't know that this is being done. In fact, there were a couple of studies cited in 2011 that suggest that only one-fifth or one-third of patients are actually getting the intervention, when in fact they already know they have one or two neuropsychiatric symptoms.

So I know I am bringing 2011 into 2009, but 2011 is sort of dead on arrival if this one doesn't go through. So I thought I had better say it now.

CO-CHAIR TIRSCHWELL: Okay. Gail?

DR. COONEY: And my comment is really just toward the group, because they are assessed -- because there are a number of

assessment measures, and I don't think we should -- I think assessment is important if we are ever going to assess interventions and, if people are not assessing, then they can't be appropriately looking at their interventions.

So I would argue that we need to look more at the specifics of the measure and whether or not they are going to be valid measures, rather than whether the -- and then use the exception criteria for those that are important and have valid measures associated.

CO-CHAIR TIRSCHWELL: So are you asking for the exception on this one, Gail?

DR. COONEY: Asking for the exception on this one? Yeah, I will ask for the exception on this one.

CO-CHAIR TIRSCHWELL: Okay. Peter?

DR. SCHMIDT: I just wanted to clarify. First, in my comment I was not specifically arguing that this measure,

everything that is addressed here, has evidence behind intervening in that case, but I think that we have had a number of situations where we have said there is no evidence for something that would be unethical to run an RCT on.

You would not get through an IRB something where you say I am going to randomize people into a cohort where they are or are not assessed for specific complaints. No one would do that. I had to negotiate with an IRB last week about how many digits of the Canadian ZIP Code it was ethical for me to collect. So they get very strict about a lot of these things.

You cannot run these things as RCTs. At the best, you can do a retrospective observational study, which is going to have a lot of confounding factors in it.

CO-CHAIR TIRSCHWELL: I guess I would say that a good retrospective study or a prospective cohort, things that are not

randomized for the reasons that you suggest, certainly fit into the multiple levels of evidence that exist and are consistently based as greater than expert opinion.

So opportunities exist outside of randomized trials to try to answer these questions.

DR. SCHMIDT: I accept that.

CO-CHAIR TIRSCHWELL: John and then David.

DR. DUDA: To that point, I think they are talking about using things like the NPI. You could easily do a trial where you do the NPI annually I a cohort and you don't do it in the other cohort, and still allow for patients to complain about their mood, and treat that.

I think that would give you some evidence on whether or not this is meaningful. But I guess I don't see how we would invoke the exemption for yesterday's Parkinson's Disease neuropsychiatric symptom measure and

not do it here. It is just a different -- It is Alzheimer's Disease neuropsychiatric symptoms. It is not just looking at memory. It is looking at mood. Right?

We know that mood is probably -- Depression and anxiety are probably poorly recognized in Alzheimer's Disease as well. I am not a dementia person, but I don't see a big difference between this one and that one, and it may very well fail for other reasons, but I think it is hard to make a compelling argument that we should have done it for the one yesterday and we shouldn't do it for this one.

CO-CHAIR TIRSCHWELL: Okay. David, are you withdrawing your comment?

DR. HACKNEY: People already said what I was going to say.

CO-CHAIR TIRSCHWELL: Okay. Bill?

DR. BARSAN: I was going to say, you know, the -- Jordan, you were mentioning about the next one that linked to this, that

if this one goes down, the next one --

CO-CHAIR TIRSCHWELL: That is not necessarily true.

DR. BARSAN: That was going to be my point. I almost see, there might even be more sense of doing the second one to show that, in fact, if you notice these things, there is a difference in terms of what happens.

Well, in this one, which seems to go one step back, at least that one is one step closer to something that actually -- where something happens, where this one is two steps away.

CO-CHAIR TIRSCHWELL: Thank you. Fred, you have one more comment?

DR. TOLIN: It sounds like we are about ready to vote on it. I will just get a final comment in as the defender of this measure.

I agree with most of everything that was said, and thank you, Jordan, for

bringing out the point again that this is paired. I think the spirit here really is we need to look at -- We need to measure it. We need to see that it is measured, and then we need to look at the intervention.

Ideally, the progression would be that at some point in time we no longer need to be looking to make sure that people are evaluating this but, in fact, looking to make sure that the intervention is being done. So I see this as a progression, and this as the first step.

You are right. There is not a lot of evidence. It is all expert opinion, but there is evidence to suggest, and it was pointed out, that this is not being done in some number of patients. So it just teleologically is a good idea to look at it and that it all makes sense, and it fits in with the intervention part that Jordan will be discussing in a little while.

CO-CHAIR TIRSCHWELL: Yes. Just

one final musing on this. I almost wonder whether a combined measure where -- You know, there are a number of ways to pass. It includes assessment and treatment. If you are assessed and there are none, you pass. If you are assessed and there are some, then you then have to move on to treatment before you can pass the measure. That might be a more comprehensive way to evaluate it and one that would sort of feel like the rubber is hitting the road a little bit better. But that would take a substantial revision from the developer.

Any further comments from the developers before we vote. Well, Michael, go ahead, and then we will go back to the developer.

DR. KAPLITT: I see the arguments for potential exception. The problem is I don't know what we are voting on here, because while this may relate more to the specification issue, I think it is important

to consider here, which is that what we are voting on is a non-exhaustive laundry list, and I don't know what it is.

I don't know what the actual measure is that we are going to be voting an exception on. We are saying it is important to look for these things, but this is not a specific thing. This is a laundry list of things that you could do anything you want, and some of it is not even on here.

So the question is: If we are going to say there is no evidence to support this but we are going to make an exception, we are going to make an exception for this -- I am not saying compelling argument -- for a thing that may not have enough evidence, but it is really important. What is it that we are even voting on here?

CO-CHAIR TIRSCHWELL: And I am not sure there is a clear answer to that. I think they have made this long list that is only a suggestion of some of the possibilities,

because there is an infinite variety of things that could come under this topic.

it is true, I think, that this is a measure that can be satisfied by documentation only, a checklist, which supposedly are not the type that are particularly preferred by the NQF.

So specifically, what we are going to vote on in a minute, we will talk about, because Gail at least has endorsed the possibility of an exception. Then one final comment from the developer before we go ahead and do the vote.

DR. JOHNSON: Yes. This is Jerry Johnson speaking. I think the case has been made for the exception by several persons who just spoke. I will just speak to this last question of what is being voted on, given the long list.

What we don't want to do here is try to specify for practitioners just which behavioral symptoms or neuropsychiatric

symptoms they have to be mindful of. The big gap -- One of the big gaps in caring for persons with dementia is just overlooking and not paying attention to these behavioral symptoms at all.

The purpose of this measure is to point out that that is a crucial part of assessment, and then that is linked to the next measure which gets to management.

CO-CHAIR TIRSCHWELL: Okay. Thank you. AMA in the room, any comments?

MS. TIERNEY: I think, just to emphasize Dr. Johnson's point and to your point about possibly a better measure, I think the intent of having the two measures is that it is exactly what you described, that you would have firstly, if there are no symptoms, then you they don't move on to the next measure. If there are symptoms, then there is an expectation that there is symptom management, and that is why the measures are paired.

CO-CHAIR TIRSCHWELL: Okay, thank you. So if we are going to move on to voting -- and somebody correct me if I get this wrong, which is possible -- if you think there is evidence by whatever standard you believe in, then vote 1. If you want to invoke the exception, I think you need to vote for number 2, and if you think there is insufficient evidence and there is no cause for the exception, you have to vote number 3. Is that correct? Okay, so let's go ahead and open up the voting for this. Go ahead and start voting now.

MS. THEBERGE: Zero, Yes; 16, No, evidence does not meet guidance; and 8 No, insufficient information submitted.

CO-CHAIR TIRSCHWELL: So because we were invoking the exception, do we then need to go on to a second vote specifically about the exception? Is that right? Okay.

I am going to just read this out loud about what we are voting on: If there is

no empirical evidence, e.g., only expert opinion, and expert opinion was systematically assessed with a group with agreement that the benefits of the process that we are talking about, which is the assessment of neuropsychiatric symptoms, that the benefit of this to patients greatly outweighs potential harms, the question is, is there an exceptional and compelling reason that the measure should be considered further?

Vote 1, Yes; or 2, No. Don't start voting yet. Please open the voting. Go ahead and start voting not.

MS. THEBERGE: Eight, Yes; 16, No.

CO-CHAIR TIRSCHWELL: So I think we are done with this measure then. Is that right? All right.

Moving on to the next measure, 2011, management of neuropsychiatric symptoms. Jordan?

DR. EISENSTOCK: Okay. We have had a lot of discussion already about this

measure, just by nature of 2009 being paired. Lots of the information is the same. Just as some background, this is taking it to the next step.

This is now the percentage of patients who had been assessed and have a known neuropsychiatric symptom, at least one, who actually received some kind of intervention.

So the numerator statement is patients who received or were recommended to receive an intervention for neuropsychiatric symptoms within a one-year period. The denominator is all patients, regardless of age, with a diagnosis of dementia who have one or more neuropsychiatric symptoms. There were no denominator exclusion noted. It is again at the clinician level. It is a process.

As far as evidence goes, it is really a redundant conversation, I think, to 2009. The same consensus arguments from the Canadian group as well as the California Work

Group were cited here.

With one exception, the comments were the same in our Work Group as well. There was some information about the quality of evidence from the Canadian group who suggested that nonpharmacologic interventions were Level 1, where pharmacologic interventions as far as quality were rated as Level 3 or expert. So there was at least one randomized controlled study to show that nonpharmacologic interventions were useful in this population.

I think that we covered everything else in the last argument. So I think maybe stopping there for comments.

CO-CHAIR TIRSCHWELL: I guess the big question here is do people think that this measure, which is closer to actions, and I guess it is my impression that there is evidence that treating these neuropsychiatric symptoms -- is that what you were referring to, Jordan? -- is supported by some evidence?

DR. EISENSTOCK: Well, the clinical practice guidelines -- that was based on evaluation --

CO-CHAIR TIRSCHWELL: It doesn't specify particularly whether they were trials or --

DR. EISENSTOCK: For nonpharmacologic interventions, and that is why there was one comment within the Work Group of maybe pairing off nonpharmacologic and pharmacologic, because of the level of evidence quality was different.

CO-CHAIR TIRSCHWELL: Okay. It sounds like there is at least some reason to consider whether the evidence isn't a little bit better in this situation than for the previous one, but I would be very open to hearing people's opinion about that. Gail?

DR. COONEY: I am never sure if I am in the right section, but my concern with this one is that it doesn't separate pharmacologic and nonpharmacologic measures,

and the evidence strongly supports nonpharmacologic measures over pharmacologic measures.

Yet in clinical practice, it is a whole lot easier to write a scrip and either one of those will qualify you for this measure. So I have a real problem with it.

CO-CHAIR TIRSCHWELL: So the potential harm is that we might be fostering something that in another measure was suggested was harmful, which is the use of the anti-psychotics or something like that. Any further comments from the committee? Gwen?

DR. BUHR: I think, just speaking about evidence, there is evidence for both pharmacologic and nonpharmacologic. There's randomized controlled trials for both, and they show that both improved. It just happens to also show that pharmacologic kills you, but it treats the symptoms, the neuropsychologic symptoms.

The number needed to treat for the

neuropsychiatric symptoms is much lower than the number needed to harm. So there is evidence.

CO-CHAIR TIRSCHWELL: Any other comments from the committee? Jordan, any other final comments before we vote? No.

Okay, let's go ahead. Daniel, go ahead. Then I have one more thing to say before we vote.

DR. LABOVITZ: I think this just is another example where we would have a lot better time figuring out whether the exception applied if there were a process by NQF to allow the developer to make that case.

CO-CHAIR TIRSCHWELL: Okay. Does anybody on the committee want to invoke the exception, which sort of changes the quality of the voting, the meaning of some of the responses? I am not seeing that here. So let's go ahead and vote.

So to move past this point, you have to vote number 1, Yes. Either of the two

or three will contribute to not passing the measure or going further. Go ahead and vote.

MS. THEBERGE: Ten, Yes; 8, No, evidence does not meet guidance; and 6, No, Insufficient.

CO-CHAIR TIRSCHWELL: So I made this mistake yesterday. I will try not to make it again. I think that means that the eight plus six is 14, which is greater than 10. So we will not move on any further with this measure. So I passed for today anyway.

Moving on to the next measure, 2016, Screening for Depressive Symptoms. Sam?

DR. FAZIO: Sure. This is dementia screening for depressive symptoms. It looks at the percentage of patients, regardless of age, with a diagnosis of dementia who were screened for depressive symptoms within a 12-month period.

They give a whole list of what those depressive symptoms could be. I am not going to read those. They are pretty

extensive, and also some examples of different scales that are commonly used in clinical practice.

The denominator statement is all patients with a diagnosis of dementia, no exclusions, and it is a process measure at the clinician level.

I guess in looking at evidence, similar to some of the discussions we have just had, this is based on some practice guidelines, but one from APA as well as the California Work Group. So no specific evidence listed and more clinical practice guidelines.

CO-CHAIR TIRSCHWELL: So they don't call out any trials where people were screened for depression more consistently and had better outcomes?

DR. FAZIO: No. They do cite -- not cite, but list the number of articles that were in all the practice guidelines.

CO-CHAIR TIRSCHWELL: That is

always a high number.

DR. FAZIO: Five hundred fifty-four for APA and 400 for the California Work Group.

CO-CHAIR TIRSCHWELL: Yes, tremendously high number. I guess it is hard to know what to do with that, in particular.

Does anybody have any knowledge outside of what was put in the document to suggest that this screening leads to better patient outcomes? salina, did you have a comment? Any comments are welcome.

DR. WADDY: I was just wondering from the developer where the depressive symptoms exactly coming from, because I was just a little confused.

CO-CHAIR TIRSCHWELL: You mean calling that out as a specific neuropsychiatric one as opposed to it being lumped or --

DR. WADDY: No, no. Like the depressive symptoms, they just put one of them

could be retardation. I assume they are talking about psychomotor retardation. Some of them are a little vague and -- Certainly, they can be associated with depression, but not specific. I was just wondering how that list came about.

MS. TIERNEY: That list, I think, is mostly based on the Cornell scale for depression. The purpose of the list really is to offer a guidance. So we realize that, as we continue in our development of measures, that a lot of users of the measures need a little bit more guidance.

I think, particularly with depression and dementia, we wanted to be explicit about what might be symptoms. So that list primarily comes from many of the elements that are included in the Cornell scale.

I don't know, Dr. Johnson, if you have anything additional to add. We did have a geriatric psychiatrist on the Work Group

specifically to work on developing this list.

DR. JOHNSON: This is Jerry Johnson. The point here is that we did not want to restrict the physicians to have to proclaim a formal diagnosis of depression, for example, a major depressive disorder which itself consists of a list of symptoms, that instead we wanted to make sure that practitioners are attentive to the fact that depression is a prevalent and potentially management problem in patients with dementia, and that they would, therefore, screen for depressive symptoms.

Whether or not they used the best ones or the formal ones that would lead to a DSM-III or IV or V diagnosis, we thought, would be too restrictive. So that is why we listed the symptoms the way we did.

CO-CHAIR TIRSCHWELL: Okay. Gwen, then John, Bill, Salina.

DR. BUHR: This one is one that I think that our discussion about the first one

applies, and what A.M.'s point was, is that I totally think that depression is often missed in people with Alzheimer's or dementia and that it is potential for the exception, because it would be important to find depression. There are good treatments for depression, and it is not easy to -- It is easy to find it if you ask for it or screen for it, but it doesn't just come up in regular conversation sometimes.

CO-CHAIR TIRSCHWELL: So you are asking for the exception. Okay, thank you very much. John?

DR. DUDA: I second the request.

CO-CHAIR TIRSCHWELL: Great. Bill?

DR. BARSAN: Some of the same things. I mean, I feel like if -- It sounds like there is good, is very common. Is there good data that treatment gives better outcome? I assume there is, but I don't know the data personally. I would be more convinced about

an exception if (a) we know it is really common; (b) we know, if you treat it, it really helps, which would maybe carry me over for an exception on detecting it.

CO-CHAIR TIRSCHWELL: Salina, did you have another comment?

DR. WADDY: Yes. I completely agree with the previous comment. It is much easier, I think, to vote for this, one, because it is so important and, two, it is not this larger grab bag, but now because of the comments of the speaker on the phone, I am just concerned in terms of how hard the evidence is for the description that you are giving for depression, since you are not talking about diagnosing people with a major depressive episode.

You are really talking about a more squishy kind of diagnosis of depressive symptoms. I am taking on Daniel's language now.

CO-CHAIR TIRSCHWELL: That is not

a good sign, I don't think. So you would be more comfortable if they said something like dementia was screened for using a validated scale or something like that, without even specifying which scale it was? Depression -- sorry, what did I say? Yes, depression.

DR. WADDY: Something less open-ended than just some of the things that you put in this list, which a lot of elderly people may have, but they aren't depressed.

CO-CHAIR TIRSCHWELL: Okay, fair enough. Hang on one second. Can anybody address Bill's question, which was supporting the fact that there is evidence that treating depression in dementia leads to better outcomes? Gwen?

DR. BUHR: I know that there are not big randomized controlled trials, but there are randomized controlled trials of people with dementia and depression who also had agitation or behavioral symptoms and were treated and improved. They are usually small

trials, but they are randomized controlled trials, and there's quite a few of them.

CO-CHAIR TIRSCHWELL: Thank you. John, you had your card up? You are okay? Any other comments? A.M.?

DR. BARRETT: I was just looking at the California guidelines with regard to this point, to Bill's question. Actually, I didn't find anything regarding treatment, but I did find a point that addresses Salina's comment about the randomized -- about using a validated tool we do have to consider.

At the expert guideline level, they cite several studies that describe difficulty in administering standardized tools to people with Alzheimer's Disease.

CO-CHAIR TIRSCHWELL: Michael?

DR. KAPLITT: I guess I have to ask -- this is more of an NQF question. I am much more enthusiastic about this than the earlier one, because it is a big problem. It is more specific, and I think the arguments

are well taken.

The problem is that, while the California argument is somewhat reasonable, Salina is right, that if you look at some of this list, the question is, if it is just going to be sort of a nonvalidated or -- as difficult as that may be, if it is going to be a nonvalidated measure that we are using, if you are going to use things like difficulty falling asleep or multiple awakenings, old men get up three times a night to go to the bathroom, you know, and that fits under this criteria.

I am not being sarcastic on the point. It is the problem with when it is sort of vague like that. We all -- I think, emotionally, it sounds like the committee is moving very close to wanting this, because depression is a very different thing than that kind of laundry list of vague, sort of affective symptoms earlier. But this is an issue.

So if overall we kind of support this, but the specifics of it, which really get more to point number two, I guess, are a problem, how do we deal with it? Do we give the developer a chance to respond and adjust it or do we say, look, you know, then that's it, and they got to come back next time with something different?

DR. BURSTIN: In some ways, I would like to try to separate what is evidence from validity, and I think what you are really bringing up is validity. So it would be good, I think, to get through the evidence, but I think your point is well taken, of can this be as precise measure in the way that it is specified? I think we will get to that. We will deal with it as soon as it gets through evidence, if it gets through evidence.

CO-CHAIR TIRSCHWELL: Okay. Peter?

DR. SCHMIDT: In the numerator statement, they talk about using valid,

reliable instruments, including but not limited to, and then a list of instruments. We should be careful about having using expert opinion to contradict a validated instrument, if we accept that these instruments are validated.

CO-CHAIR TIRSCHWELL: I am sure they are validated in the general population. Whether they are specifically validated in dementia, I guess, is the issue at hand, and you are right. It is already in there. So I don't know why we would necessarily have to add it. Maybe we have to make the wording more strongly that they need to do it or something. Salina?

DR. WADDY: Yes. I definitely saw these, and I saw that the Cornell Scale for depression, which was actually in dementia was there, but that really -- That gets back to the question that I asked the person who is on the phone, as well as the developer, you know, how they came up with these. If they said

that all of these symptoms came from these three scales and how they were used, then I am fine with that. But that is really not what they said.

CO-CHAIR TIRSCHWELL: Ramon, and then John.

DR. R. BAUTISTA: What about the patients with severe dementia? Can we really assess them for depression? These people are pretty much nonverbal at that point. How do you account for them?

CO-CHAIR TIRSCHWELL: Okay. John?

DR. DUDA: I just want to request that we vote on the evidence, because I think these arguments are going to play too much into that decision, and I don't think it is appropriate.

CO-CHAIR TIRSCHWELL: That's fine. Let's go ahead and vote. I think at least one person has called out -- More than one person has called out the exception. So when we vote on the evidence, again number 1, there is

evidence to proceed; number 2 means you are in support of the exception; and number 3 means you think neither is there evidence nor should the exception be applied here. So let's go ahead and vote, starting now.

MS. THEBERGE: One, Yes; 19, No, evidence does not meet guidance; and four, No, insufficient.

CO-CHAIR TIRSCHWELL: So now we will move on to the voting on the exception, if you could throw up that slide. John, do you have a point first?

DR. DUDA: Maybe I should have done this with the other one, but I would like to hear some argument. I was kind of expecting the other one to get the exception as well. I would like to hear the argument against the exception for this one, if people are planning on doing that.

I think it is an easy argument to make that there should be an exception, that this is a compelling and, whatever the wording

it, exceptional case where there is little chance of harming a patient if you ask them about depressive symptoms, and a compelling reason that it might be helping them.

CO-CHAIR TIRSCHWELL: Exceptional and compelling reason. David, did you have a comment? Then Daniel.

DR. CO-CHAIR KNOWLTON: I can't give an argument one way or the other, John, but I can say that I am troubled with the exception in general. I think that we forget that NQF has a whole bunch of committees doing a whole bunch of things that are making a whole bunch of providers go through all kinds of hoops, and I think the standards for those should be very high.

I am a quality advocate. That is my day job. That is what I do. So I am in favor of quality, and I want us to kick butt with quality. But I feel that we already have providers beside themselves with stuff they have to report, that they have to comply with.

Part of what I do is defend it by saying this is a high standard, harmonized measure, and I was part of the committee that worked on harmonization with Leapfrog and NQF and feeling strongly that you can't be jumping through everybody's different hoop.

What happens when provider groups come in? They don't want to be prescriptive. They really don't. They want to say, well, I got to refer to another doctor, and NQF doesn't let them do that. Then they say, we don't want to impose upon clinical judgment. NQF doesn't want to let them do that. They want to say you can say professional judgment has a place, but you have to document it, and you have to make other decisions.

So it is trying to make it robust. I can't speak to this particular measure, because I am not the clinician here, but I do get troubled with this exception, and it is has been troubling with every single vote, same issue. I guess, but on the other side of

that argument, you guys are the clinicians, and I say you want to go through those hoops, that's okay with me. But it seems to me that we are trying to get standards that we can hold people accountable.

When we look at making the exception, we are talking about it being exceptional and compelling. I understand that. Peter made that argument during the break, and he is right, but sometimes they are just compelling arguments.

I believe we were talking about the epilepsy and pregnancy where it is sort of like the HIV argument we talked about and so forth. You know, safe sex was just a good idea. You could think about it logically, and maybe we haven't got the measures yet. We need to get them, and so forth. I get all that, but I think that -- I think this should be a very high jump bar.

It should be a very high bar to have somebody get an exception. I was

troubled the last time with somebody coming in, and I'm with John. I thought we were going to debate it, and we went right to a vote.

I was troubled in the last one that the person presenting the evidence said I think an exception would be appropriate, and I am uncomfortable with that.

So that is not an argument for or against. It is just a statement of how I feel about the whole exception thing.

CO-CHAIR TIRSCHWELL: A word of caution that there are costs associated with all these things. Daniel?

CO-CHAIR KNOWLTON: Absolutely.

DR. LABOVITZ: I heartily second Dr. Knowlton's comments here. As one of the guys who has to actually do these measures and flog my people to do them as well, there is a very heavy penalty for adding another set of checkboxes that have to be done when you are trying to evaluate a patient.

I don't find this compelling, and here is why. It is not that I don't care about depression in demented patients, because I deeply do. I think depression has a huge impact on quality of life in almost every disease where it is more common than in the general population. Often it is more important to treat the depression than it is to treat the disease. In epilepsy, I wrote a paper with my wife that seemed to show that.

In this case, we are not talking about depression. We are talking about depressive symptoms. We are removed from the remove. I don't find that compelling.

CO-CHAIR TIRSCHWELL: John?

DR. DUDA: I think, David, you were making the argument that it was not exceptional and compelling, and I accept that. I think that, Daniel, in your practice, you know, the argument that we are going to be making people in practice centers jump through more hoops -- I don't know if that flies with

me, because -- and I certainly check for depression in my patients. So I am not going to have to do anything better.

I thought the purpose of this was to kind of assess whether or not people are doing standard of care health care practice for the benefit of these patients, and I think that assessing depression in a dementia patient is a no-brainer, standard of care aspect of caring for those patients.

So people who are doing good care are not going to have to do anything different, and to say -- You know, we are not arguing about whether this particular measure is the right way to do that. I think at this point, we are arguing about whether or not the lack of empirical evidence to prove that this should be a topic that has a measure about it is worthwhile or not.

This may very well fall on the validity/reliability issues or usability, but is this an exceptional or compelling reason to

step past the need for a compelling empirical evidence

CO-CHAIR TIRSCHWELL: Peter?

DR, SCHMIDT: There are a number of cases where things that seem obvious and they are important to do wind up having negative consequences. So Ramon brought up the severely demented patient. How are you really going to assess them? But if you set this as a standard, people will carve out time and do something.

One of the complaints that I have gotten as I have gone around and talked to movement sort of people is that in meaningful use, they said smoking cessation is required for meaningful use.

In Parkinson's Disease, there is some evidence that you shouldn't be counseling people to stop smoking. It is probably not what is going to kill them.

Some people feel that they are self-medicating by smoking, and it is very

frustrating to these people that they are required to counsel these people about smoking when, in fact, a lot of these people have decided that smoking is beneficial to them, to their Parkinson's Disease. So that was adopted as one of 10 things that was put into meaningful use, and it is a negative in this case.

To your point, are we going to wind up with people who are wasting their time assessing people for depressive symptoms when they are severely demented and can't respond?

CO-CHAIR TIRSCHWELL: Salina?

DR. WADDY: All of those points are certainly well taken, and I think I agree most, though, with what Daniel was saying. Is this measure really getting at the symptoms that -- or the problem that we are trying to go after? I think they just throw a lot of things into that. I am not confident that it really reaches a level for the exception.

The other thing: Since I have

been here and I have access to the Internet and this has been coming up, I am a little bit troubled that some of the things I am finding regarding the geriatric depression scale and the attempts to use that scale, because you specifically listed it -- use that scale in patients with dementia.

There have been studies that have shown that it really loses its validity when you try to apply it. So can you talk a little bit to that?

MS. TIERNEY: Sure. I think the overall intent of the measure is not to be prescriptive in the manner in which the clinician would screen for depression.

I think, you know, this measure is unique compared to other screening for depression measures in that it specifically we are screening for depressive symptoms, but it is actually in many ways very similar to other screening for depression measures that have been endorsed by NQF and that are out there,

in that you are not making a diagnosis of depression. You are simply screening for symptoms of depression, and any of the validated tools potentially do that.

Then the next step is actually to do a formal diagnostic examination. So I think this measure is actually not congruent with any other screening for depression measures. It might be more apparent to specifically use the term screening for depressive symptoms, but certainly any type of screening measure, you are simply doing a screen and a quick check to see whether or not more evaluation might be needed. Then more evaluation can be done to make the formal diagnosis and then ultimately some type of treatment.

So I think that is the spirit of this measure. I think it might be difficult with the terminology used to answer your specific question. I think, as our Work Group discussed, the ideal would be that Cornell

scale for depression and dementia, but as we hear from -- As many of you have noted about the difficulty in using these in practice, if you get too prescriptive, then maybe the -- and I am not sure how long -- Dr. Johnson might be able to compare.

So if we were to put this in practice and say we require the Cornell Scale, we would hear a lot of people saying I don't have 20 minutes to administer the scale, but that is why we do allow clinical judgment, and we have listed other scales that are also available for certain patients.

For specifically for geriatrics, PH-29 is a very broadly used scale, and I guess clinicians might see most appropriate for patients needs and maybe the confines of their practice setting.

DR. WADDY: Regarding the GDS and that it remains valid in older patients, but actually applying it to the demented patients, which is a good measure.

CO-CHAIR TIRSCHWELL: so we should probably come back to that if we get to validity. Ramon?

DR. R. BAUTISTA: I am quite concerned to actually discussing an issue or a measure where the developer, as he says, the quality of the body of the evidence was not addressed. You know, this is going to be one of many, many things we are going to require physicians, clinicians to do.

If you have worked in compliance committees before, trying to justify every single clinical note, you know how it is sometimes to get everything right. It sounds like a good idea on face value, but the point is -- I mean, I could at least ask for a case series to justify whether you want this measure to come about, and there is nothing here at all even close to that.

I am just concerned, like what David said. I mean, we are actually asking more and more of physicians, force another

test on the nurse practitioner to perform on patients, and for really what is the evidence for that? You know, at least show me a case here that this actually works, you know.

CO-CHAIR TIRSCHWELL: John, Risha, then Gwen.

DR. DUDA: I guess I just wanted to say again that I think a lot of this is related to validity and reliability, and I haven't heard any compelling reason other than you just don't believe it is exceptional and compelling, that they looked at the literature. They said that it is not -- it was systematically assessed.

Is there an agreement that assessing depression in demented patients -- and don't talk about how they are assessing; I don't think that is relevant here -- greatly outweighs the potential benefit, greatly outweighs the potential harm? I think that is the question on the table. Right?

CO-CHAIR TIRSCHWELL: Yes.

Depressive symptoms, not depression. Gwen, then Risha.

DR. COONEY: I agree with John. I think that a lot of the literature on depression and dementia is more about depressive symptoms. So because of the difficulty of making a diagnosis of major depressive disorder, they don't talk about that. They talk about depressive symptoms, and do you have dysphoria, and then you treat that, and then they get better.

So I really think that, in contrast to the previous one, that the potential benefit does greatly outweigh the potential harm.

CO-CHAIR TIRSCHWELL: Okay. Thank you. Risha, then Michael.

DR. GIDWANI: what I see as a potential benefit is that a patient that has been diagnosed with depression getting linked to treatment, that that improves their outcome; and I don't think we are there with

this measure.

We are looking at depressive symptoms. Then there has to be a formal evaluation of depression. Then we are hoping that the patient is going to be linked to the appropriate management, and I think we are now getting really far away from the health outcome of reducing depression that we are interested in.

CO-CHAIR TIRSCHWELL: Michael and Daniel.

DR. KAPLITT: To David's point about standards, I completely agree. As much as I want to support measures like this, I would kind of turn the argument around and say, rather than what is the compelling argument not to invoke the exception, I would like to know what the compelling argument was why evidence was not provided here.

We are not talking about randomized, double-blind studies. We cannot generate evidence or provide evidence in this

document from everything that has done out there in the world that says that evaluating and treating patients with depression actually has benefit in any reasonable form? We are giving expert opinion, and then we are saying, well, we don't really have access to how they did their data, because they don't provide that online, whatever.

We are lowering our standard to a level here. This is not a data gathering organization, and this is not sort of a think tank. To the point of, well, there is really no harm, and if you are doing it anyway -- for every couple of minutes that I spend on each of my patients, surgical patients, having to document and discuss body mass index and smoking, when you add that up throughout the day, that is easily two or three more patients that I cannot see because of those things that I am doing, and those patients that I have no relationship with could care less about my discussion with them about their body mass or

their smoking, but I have to do it, because it is required, even though it does nothing.

So there are costs to all of this unless we have a real standard.

CO-CHAIR TIRSCHWELL: So, Michael, you are sort of bringing out the point that --

DR. KAPLITT: I would like to know why we are even invoking the inception, because while I am emotionally very supportive of this idea, I would like to know why we don't have better evidence here; because, to me, part of the reason for the exception, whether it is written or unsaid, is that there is just no good way to get any type of compelling evidence. Forget about randomized, double-blind studies.

CO-CHAIR TIRSCHWELL: And you don't believe that is the case here?

DR. KAPLITT: I don't see how that is not the case here.

CO-CHAIR TIRSCHWELL: Okay. Very good. Daniel?

DR. LABOVITZ: I am back to beating dead horses. I think part of the reason that we are struggling here a bit is that each one of us individually has invented his own internal set of standards for what exceptional and compelling is.

When we have evidence based measures, we go through a lengthy multi-step subsection 2, Part 17. It is extraordinary. Here it is just like, hey, what do you think? Is this big? I can see why there is disagreement in the room, and I think we have been asked to do something that we are not prepared to do, and that developers weren't set up to present.

I think this is really a demand. The struggle we are having here is really a demand for, if we are going to have the capacity to have exceptions, we need to have a set of standards and a set of agreed upon rules for how to approach it.

I think John Duda and I completely

agree that talking about depressive symptoms to patients is a fundamental part of clinical practice. Where we might disagree is whether or not you go through the rigmarole of checking off 17 boxes in an EMR form to document it, and whether you then go through the rigmarole of measuring that, because you know it makes a difference to do the task.

That is a fair disagreement, but we don't even have a basis for having that disagreement. All we can do is to say, ah, I think it is compelling or I don't.

CO-CHAIR TIRSCHWELL: Salina?

DR. WADDY: I agree with Daniel's statement, but I also agree with Gwendolen and John in that this is a very important topic for them to develop a measure. The issue that I have is that exceptional isn't regarding whether or not the topic is important. It is whether or not this measure meets that level, and I just don't think so.

CO-CHAIR TIRSCHWELL: I think that

also gets back to Risha's point. Is it so far removed, as has come up with many other measures, that it is hard to have that faith that it is going to translate into better outcomes?

Any other comments that people have? Yes, Helen?

DR. BURSTIN: Just one brief comment. When our Evidence Task Force did this work about a year and a half ago or so, this was really intended to be an exception. They didn't spend a lot of time on it, because they really focused in on the fact that they wanted to see quality, quantity, and consistency of evidence. But then, as the discussion really emerged, there were clearly areas in clinical care and health systems improvement where evidence was emerging in some really important topic areas that just may not be there yet.

So this was not to be an exception for when the developer couldn't cull the data

and put it forward. It was really an exception to when the evidence just wasn't there yet, but it was such an important area that people thought it was important enough to bring it forward for now.

I think your point is well taken. If we are starting to see committees struggling with this and trying to invoke it more often, we need to go back and standardize exactly what we expect of you guys, what we expect of the developers. But again, it was called an exception intentionally to be a rare event, not something that we just reflexively go to if the developer can't provide evidence.

CO-CHAIR TIRSCHWELL: And just as a point of clarification, if the guidelines that are out there don't spell out the evidence in enough detail, does it then become the obligation of the developer to go back and look more at the primary literature to be able to present it themselves?

DR. BURSTIN: Right, and that is

the intent. For those of you who didn't see it, the IOM came out with a report fairly recently on the quality of guidelines in America. So we are sort of in a supply chain, of course.

So if the guidelines aren't doing a good job of providing transparent systematic reviews on the quality, quantity, and consistency -- and those are the exact words in the IOM report. They said guidelines should be clear as to the quality, quantity, consistency of the evidence.

So if there is a systematic review done, that's great. they can cite the systematic review. But if that is not transparent, it is a burden to the developers, and we understand that, and a lot of the developers aren't set up to certainly do their own systematic reviews, but it is, I think, something we will see change over time as guidelines improve.

CO-CHAIR KNOWLTON: But, Helen, I

think that getting some level of feedback is actually key to transparency. That is the problem here. I agree with Daniel completely. You know, it is everybody's kind of seat of the pants judgment here.

I don't disagree with what John or what anybody has said, Gwen or anybody is saying this clinically. I am just saying that we haven't got any criteria here. It is Dave's criteria and Dan's criteria and John's criteria. That defeats transparency.

So I think -- and as we said when the evidence committee came up with this exceptional thing, maybe they should have given us an example of an extreme. We have something, and we say does it rise to that level, because they could, as Michael said.

I think the issue here is -- I don't disagree with what John and Gwen are saying. This is important in clinical practice. As Salina said, let's go get the evidence. There's got to be evidence out

there.

DR. BURSTIN: And the intent as well of making it a very transparent exception -- again, you guys are still early in the process. When this goes out for comment, it is obvious to everybody out there reading this report, this measure went forward on the exception, and then we get comment on that.

So while it is somewhat reliant on the perspective of the multi-stakeholders sitting at this table, it then is fully transparent and goes out for broader public comment to get a sense of was that exception reasonable.

CO-CHAIR TIRSCHWELL: Jolynn, then Ramon and John.

MS. SUKO: Well, it sounds like that this is really evolving, and I am thinking that there have been a number of suggestions around the table, like Peter's suggestions. Is it an area where it would be unethical to do research?

I am wondering if the NQF can cull some of those from the conversation at this table as it matures and moves forward to better define some of this.

CO-CHAIR TIRSCHWELL: Ramon?

DR. R. BAUTISTA: If there is no evidence -- There is an evidence based discussion. I believe it is the developer's responsibility to tell us why there is no evidence. I mean, they should tell us that explicitly, there is no evidence because, not just to leave it hanging like this.

CO-CHAIR TIRSCHWELL: John?

DR. DUDA: I agree with that, but we are here now, and I think it is -- we are not going to get them to change their guidelines or specify exactly what they mean, and I think in the interest of transparency, we have to remember that yesterday we made an exception to a measure evaluating depression and a bunch of other things in just a squishy way for Parkinson's Disease, and now we are

saying that dementia patients -- that it doesn't meet that standard.

CO-CHAIR TIRSCHWELL: I think that speaks to the lack of standards in making this decision more than anything else, and that it is probably impossible to be consistent without standards. It is definitely a seat of your pants thing, and some measures people, people think, are more important than others.

Ramon, do you have a final comment or can I -- So let's go ahead and vote on the exception here. If you think there is an exceptional and compelling reason, vote Yes; if not, vote No.

MS. THEBERGE: Six, Yes; 18, No.

CO-CHAIR TIRSCHWELL: Okay. I guess we are done with that measure. Which one was that anyway -- 2016.

All right. Last one before break, 1990. Daniel, can you lead us through an overview and the evidence, such as it exists?

DR. LABOVITZ: This is a measure

of grading severity of dementia in patients with an ICD-9 code documenting the presence of dementia. So the numerator is whether a severity of dementia was -- whether dementia was classified as mild, moderate or severe at least within a 12-month period, amongst all patients with a diagnosis of dementia.

it is strongly implied but not strictly stated that one of a number of available valid and reliable instruments should be used to assess the dementia.

This is a patient level -- a provider level measure, and it has exactly the same amount of evidence cited as all the previous measures.

I did take the time to look at the six citations mentioned by the developers. I didn't want to spend that time, but I was curious to see the six reference out of the 556 that are in the American Psychiatry Association Guidelines.

Each one of those references is

simply a reference to one of the tools that one might use. It has nothing to do with evidence for the measure. I would like to see down the line -- a little editorial here -- that -- Just spare me that. I don't want to have to look to see that there is nothing there. I think the developers perfectly knew that. We all did.

The question is going to revolve here, as it has with all the other measures on dementia, is there a compelling exception, and that was a major focus in the committee discussion.

CO-CHAIR TIRSCHWELL: Mary, then Gail.

MS. VAN DE KAMP: I just have more of an information question. It is already a PQRS dementia measure. So are they already -- Are physicians already reporting this as part of that measurement system? Does anyone know?

MS. TIERNEY: Yes. In 2012 not all of these measures presented for you are

part of the dementia measurement which is currently being reported.

MS. VAN DE KAMP: So is there data then now that you are collecting in the PQRS information?

MS. TIERNEY: Yes. CMS is currently information on these measures. It usually takes some time for that to make its way to us so that we can get that information and see how physicians are performing on the measures.

We do receive some patient comments. So if someone had a question about the measure and how it is supposed to be used, those often are directed to the developer, and to my knowledge we haven't really received anything on any issues of concern with these measures.

They are currently being used in that program, but it takes a year, if not longer, in order for us to get some data from CMS on these measures.

MS. VAN DE KAMP: So that means there is a financial incentive for physicians to use these outcome measures. Is that correct? with the PQRS?

MS. TIERNEY: Yes, there is an incentive payment. Yes.

CO-CHAIR TIRSCHWELL: You don't use them.

MS. VAN DE KAMP: You don't get money forward. You get money back? There is no disincentive.

DR. BURSTIN: No. They take away your money. Currently, it is still an incentive program, but in 2015 penalties will start.

MS. VAN DE KAMP: I guess that is one of my confusions with some of these that are PQRS measures and NQF measures. Is there any sort of harmonization with CMS to these? I mean, do they look at NQF to come back to say --

DR. BURSTIN: Yes. The majority

of the measures on PQRS are NQF endorsed. Some of these newer ones were put on the PQRS list in advance are reviewed by NQF.

CO-CHAIR TIRSCHWELL: and I think in our first round through all these dementia measures, we were confused by the fact that they were already in use, but there was no data. Then the developer came back to us and said, yes, they are starting to be used, but we don't have the data yet; so we can't present the data. Gail?

DR. COONEY: I was waiting for Daniel to invoke the squishy clause on this one. My problem with this one is that I could find nothing anywhere in the evidence submitted to support consistent division of dementia into mild, moderate and severe categories, and I don't see how you could measure something when there is inconsistent guidance on what they are.

CO-CHAIR TIRSCHWELL: A.M.

DR. BARRETT: I will just briefly add to what has been said, that I have found some evidence that may be of interest to the group, and it was actually in service of the next measure, but also applies to staging.

I would urge, as Daniel has said, developers to be guided to produce this kind of information as part of the application, because that would be very helpful.

First of all, the data is only in a research setting, and you have to make several leaps to apply this data, but it can be said to apply.

For example, the first study regards the outcomes of patients taking part in Alzheimer's studies who did or did not have cognitive assessment. So there are many confounds, obviously.

In a study in Australia 1900 patients were evaluated initially for a research study, treatment research studies, and 246 did not complete the evaluation to the

point of cognitive assessment, and those patients had worse outcomes.

The other evidence is even softer than that, unfortunately, but there are two other studies of reports by caregivers, a small study of benefit of participation in clinical research reported by caregivers and patients which stated that assessment is one of the benefits that they perceive to be useful and appropriate for taking part in research; also a study of behaviors, physician behaviors, that lead to referral for clinical trial participation. So, again like 19 leaps you have to make there for outcome, but essentially, obviously, people are more likely to be referred for a clinical trial if physicians have access to diagnostic instruments and apply them.

Lastly, I would just say that in this instance, staging will help to differentiate between mild cognitive impairment and mild Alzheimer's Disease, and

there is probably both some public health and individual patient benefit on that regard.

CO-CHAIR TIRSCHWELL: Although they have to have a diagnosis of dementia to even qualify for this. So you are suggesting it will lead to some diagnostic reclassification? I see. Sam, and then Risha.

DR. FAZIO: I guess I would just like to add some anecdotal comments. We hear from families all the time that the various classification systems are confusing for people, because there are so many different ways to classify stages.

A consistent way to stage people or to group all these scales in similar type stages can help give people sort of a system to better make decisions about care and also to sort of deal with what might be happening in sort of that vague stage or that larger, broader stage instead of very specific stages.

Everybody doesn't fit into these

little boxes sometimes that these scales sort of put people into. So having these three larger staged gives people a little bit more variability.

At the same time, I think you see when people are staged incorrectly how that leads to all sorts of labels and inappropriate care, poor quality care, and inappropriate expectations of what is going to come. So I think a system that sort of groups all these scales that are out there into some broader systems, I think, would be really helpful for families and people with disease.

CO-CHAIR TIRSCHWELL: Does this measure get to that, or not?

DR. FAZIO: Yes.

CO-CHAIR TIRSCHWELL: They do? So that was a thumbs up type of comment?

DR. FAZIO: Yes.

CO-CHAIR TIRSCHWELL: Okay. Thank you. Risha.

DR. GIDWANI: I read the cognitive

assessment, a functional status assessment, first. Then I read the staging of dementia. It wasn't clear to me from reading the developer's report for staging of dementia what this is going to give us in terms of being able to better hone in treatment practices for patients that the cognitive assessment and the functional status will not provide.

I am hoping maybe the clinician experts in the room can elucidate me on this regard.

CO-CHAIR TIRSCHWELL: Anybody want to respond to Risha's request? A.M.?

DR. BARRETT: There are specific indications for treatment, for example, cholinesterase inhibitors or other treatments, either recommended by research, manufacturer or third parties that refer to stages rather than to specific scores on cognitive assessment.

CO-CHAIR TIRSCHWELL: Salina?

DR. WADDY: I don't understand them on that. I completely understand that there is a need to stratify degrees of dementia, but has that already been agreed upon like by the Alzheimer's Association or by thought leaders in terms of taking all of these different tests and agreeing upon what is severe, mild and moderate; because otherwise, it seems like that would -- If that hasn't been agreed upon by those thought leaders, it seems like it is a large step forward, backward or sideways to come here and ask us to really push that forward.

So I am not an Alzheimer's expert, but your comment confused me a little bit.

DR. FAZIO: Sure. Well, we do use mild, moderate, and severe at the Alzheimer's Association, but we haven't looked at these assessments and grouped them that way.

DR. WADDY: Well, that is what I am saying. This is a big step. It seems like this takes us --

DR. FAZIO: But I guess my assumption was that their group of experts that came up with the clinical guidelines would have done that.

DR. WADDY: I don't know.

CO-CHAIR TIRSCHWELL: I think that is part of everybody's problem, is that we have to make all these assumptions, because a lot of it is not spelled out in the application set or filled out. Jane?

MS. SULLIVAN: I would agree with Salina, and I agree with you that, when we are talking the same language, it is helpful not only for practitioners but certainly for families.

The way I read this, there are six or seven different scales that are suggested, but I don't read that there is any consensus about the way in which people would be assessed and the way in which it would be staged. So I don't think -- I don't read that it addresses the point that you are raising,

which I think is a really valid point.

CO-CHAIR TIRSCHWELL: Any other comments before we -- Yes, Ramon?

DR. R. BAUTISTA: Yes. I am concerned about the statement that says the quality of the evidence was not addressed. Again, as a committee we are not really here to provide the evidence. We are here to assess the evidence, so we might know that it might be good for this or that reason, but the fact is it is not presented as an evidence to review.

CO-CHAIR TIRSCHWELL: Okay. Thank you. Before we vote on this evidence, does anybody specifically want to invoke the exception for this measure? I am not seeing any response. So then as we are voting for this, the only way to move forward is to vote: 1 as Yes; either 2 or 3 would be a vote to not move forward with any further evaluation of this measure. Let's go ahead and start the voting now.

MS. THEBERGE: Zero, Yes; 10 No, evidence does not meet guidance; and 14, No, insufficient information submitted.

CO-CHAIR TIRSCHWELL: So we are done with this measure. I think that brings us to our break. We are just a little bit behind schedule, not bad. so let's take a 15-minute break, and reconvene at ten minutes before 11. Thank you, everybody.

(Whereupon, the above-entitled matter went off the record at 10:34 a.m. and resumed at 10:52 a.m.)

CO-CHAIR KNOWLTON: We are going on to -- let me see -- 2000 Dementia: Cognitive Assessment. Dr. Barrett.

DR. BARRETT: Welcome back from the break, everybody.

In this measure we have much to say and issues that had come up previously. In fact, as you heard, I kind of brought out my little carpetbag of evidence one measure early for Daniel's presentation. But as we

are considering the evidence, of course, cognitive assessment is part of a clinical practice standard in the assessment of dementia.

I think, in the initial assessment of dementia, that many people would appreciate the potential for patient harm in misdiagnosis of Alzheimer Disease, either positive misapplication of the diagnosis to people who have ALS, early PD without motor symptoms, even brain tumors, of course, depression as we have talked about, but even these rare disorders like atypical dementias, epilepsy, B12 deficiency, and once I saw someone with a factitious disorder who, of course, had been diagnosed with dementia previously, infectious diseases like HIV and neurosyphilis.

Now nobody is saying that those account for a large number of people, of course, and then also it would be very difficult ever to do any kind of a prospective study on this topic, because even if we were

able to look and -- Well, notoriously, looking at dementia, looking at the application of a diagnosis correctly versus incorrectly is very difficult to do.

So let's focus on the second rationale that we talked about before of targeting treatment, and I will simply present that in this instance we have a different situation than staging in that in cognitive assessment we can identify the profile and the specific symptoms.

I believe that the APA guidelines specify four areas. I can't cite them for you right now, but I think it is like visual spatial function, memory, attention, etcetera, and there are people who have variance in the syndrome.

So large numbers of people with dementia may, for example, have a lot of behavioral symptoms, a lot of language symptoms, but not much in the other areas, and they do require specific treatments. It goes

the other way, too.

So somebody may have aphasia from Alzheimer Disease, may have a lot of behavioral disturbance, but may be able to draw a beautiful correct clock, and that person may, in fact, be relatively functional. So the history, of course, is misleading in these people.

Again, as we have commented, the opportunity for specific evidence from, for example, care records was not taken advantage of in this application. So we don't have presented, for example, evidence from CMS records that hospitalizations or other secondary visits may be less in people who received cognitive assessment, and there is a CPT Code for cognitive assessment, actually several of them.

So this is disappointing that we don't have that kind of information to consider, because as was commented in our Work Group call, given that cognitive assessment is

more common probably in a subspecialty setting now, that in the general practice community it is likely that that would potentially support the benefit of applying this criterion or this standard.

I already listed for you evidence from a research setting that cognitive assessment may be beneficial. You have to believe, if you look at those studies, though, that it was the cognitive assessment piece specifically rather than other aspects of research participation. The Australian study, as I said, of about 2,000 patients does somewhat support this, although it is confounded by dropout.

Unfortunately, the developers only presented for us consensus measures from the APA, and I think they make reference to another consensus measure from California. So we have that.

Questions that would arise with respect to the rationale include initial

versus repeated assessment of patients. So it is really important to do an annual evaluation? Again I would say that all three of the reasons, initial diagnosis, targeted treatment, and public health, would all apply. The same argument would all apply but, of course, it is a smaller group of people.

The last argument I think you can make about this is that consistently there is an argument about the value of the clinical trial data that we have in Alzheimer Disease. People always talk about how people who take part in clinical trials are different from the typical population.

Of course, they probably underrepresent disadvantaged groups, but also we have to consider that there may be some benefit again of the cognitive assessment that these people receive as part of their clinical trial participation. So it may not be that there is a selection bias, but it may actually be a "nonspecific" treatment effect.

I tried to find -- There was a study, I think, published in Neurology in the Nineties that supported the idea that people who take part in clinical trials just do better, whether they are in the placebo arm or whatever. Unfortunately, I couldn't find that study for today.

Again, unfortunately, all we have from the developers with respect to evidence is clinical practice guideline evidence. I was unable in reviewing those guidelines to find specific studies that support cognitive assessment, and this is also one of these PQR measures, PQR measures that are being used in a trial period. I believe this is up for trial one year endorsement. Is that correct?

CO-CHAIR KNOWLTON: Questions? David? Anybody? Okay.

CO-CHAIR TIRSCHWELL: I guess I would just add that A.M. has done extra work here to try to identify an evidence base, which really wasn't presented by the

developers. In reality, it is more the same as all the other measures which didn't do well than different, is my perspective.

CO-CHAIR TIRSCHWELL: Let's vote. This is on the evidence.

MS. THEBERGE: Zero, Yes: No, evidence does not meet guidance, 9 votes; and No, insufficient information, 15 votes.

CO-CHAIR KNOWLTON: That's it. David on 2004, Functional Status Assessment.

CO-CHAIR TIRSCHWELL: Okay. So this is another dementia measure. Just to review what the measure is, the description is the percentage of patients, regardless of age, with a diagnosis of dementia for whom an assessment of functional status is performed and the results reviews at least once within a 12 month period.

The numerator statement is those for whom a functional status -- an assessment of functional status is performed, and they give some examples of that, including some

scales, and the results reviewed. The denominator is all patients, regardless of age, with a diagnosis of dementia.

There is an exclusion for a documented medical reason for not assessing functional status. There is no risk adjustment. It is at the clinician level, and again this is one of the ones that I think is being used in PQRS. At this point, though, no data were presented.

Then focusing on the evidence, the quality, quantity -- excuse me. Starting with the quantity, they do refer to a number of articles in the consensus papers, but those are just some numbers. The quality, they really don't describe at all. In fact, it is not in the consensus statements, and no further work was done. As far as consistency, that is not commented on in the guidelines either.

I go back to a quotation from one of the guidelines, which says, and I will read

this -- it is a little bit long: A detailed assessment of functional status may also aid the clinician in documenting and tracking changes over time, as well as providing guidance to patient and caregivers.

Then they describe what functional status might be: These regular assessments of recent cognitive and functional status provide a baseline for assessing the effect of any intervention, and they improve the recognition and treatment of acute problems such as delirium.

So that is sort of the rationale in connecting the assessment with a hopeful good outcome, but there is no real evidence to support that connection, and the particular guideline says that this recommendation statement was not even rated, which is -- You know, they have some that are rated with high confidence, and I realize that is yet another rating scale that we are not so familiar with, but this one isn't even on their rating scale.

They didn't rate this thing.

So I think in many ways it is similar to these other measures, not a lot of evidence. Virtually none was presented by the developers.

CO-CHAIR KNOWLTON: Any questions for David? Okay, on the evidence, let's vote.

MS. THEBERGE: I need one more.

Zero, Yes; No, evidence does not meet guidance, 11; and No, insufficient information, 13.

CO-CHAIR KNOWLTON: We are on a streak. This is 2028, counseling regarding safety concerns.

DR. RICHMOND: All right. Counseling regarding safety concerns. This is actually one step closer to outcome, because we are beyond assessment and now to counseling.

It is looking at the percentage of patients, regardless of age, with a diagnosis of dementia or their caregiver, who are

counseled or referred for counseling regarding safety concerns within a 12-month period.

So the numerator are patients who are counseled or referred, and counseling is defined. The denominator is all patients, regardless of age, with a diagnosis of dementia. There are some exclusions, which is documentation for medical reasons, for example, end of life or other medical reasons.

Jumping right to the evidence, this really does have the same issues as the previous, but I would say as a nurse and an injury scientist, counseling is really appealing to me.

So I looked, and there was evidence, I thought, although I don't think it is really true, showing dementia, increased risk of falls or wandering, then injury and death. But actually, the citations supporting that structure process/outcome link of just does this happen was actually an instrument development study that was looking at

interrater reliability. So it really was not evidence even showing that provided.

Then looking for either the quantity, quality or consistency of evidence showing, if I counsel, does that improve outcomes, it has the same issues as before, as the evidence really was not provided. So the Work Group had significant concerns about that.

CO-CHAIR KNOWLTON: Questions or comments on the evidence? Okay, we will vote on this one.

MS. THEBERGE: One Yes; 10, No, evidence does not meet guidance; and 13, No, insufficient information.

CO-CHAIR KNOWLTON: The next one I am presenting, which is on counseling regarding the risks of driving. A couple of items on this one.

First off, I was going to say, if Terri's went through, I was going to say it is already contained in the one that she did. So

it was completely duplicative, but I believe, and the Work Group felt, it has -- I won't go through it all again -- exactly the same concerns that Terri raised in the first one. They sort of cut and pasted the same presentation, the same type of information.

So there was no reliability or validity data provided or any indication that this was making a difference. Again as Terri outlined, I think people intuitively felt that this was something you want to do, but there was just no evidence or any way to comply with the measure in a statistically or a consistent -- I didn't want statistic there -- consistent fashion and know that you are doing it consistently.

Questions or comments? Okay. Oh, wait a minute. We do. yes?

DR. BARRETT: I spent a lot of time thinking about whether I would want to request an exception on this measure because of the -- My personal feeling about how

compelling this measure is, is comparable, I think, to risk related to pregnancy and to epileptic drugs, driving with seizures for example.

I think the reason why I came down finally with a no was that it is a counseling measure. So I just wanted to --

CO-CHAIR KNOWLTON: It is a counseling measure. So you felt, if it were an assessment or -- What are you saying? If it were an assessment? Just finish the thought.

DR. BARRETT: Well, I hate to make recommendations to design the measure myself, but if it were closer to an intervention or assessment exactly, I would be more enthusiastic.

CO-CHAIR KNOWLTON: The reason I pushed you on it is because there is a transcript, and so developers will listen, and that is the important thing. So that is why I asked you. Salina?

DR. WADDY: This measure and the last measure are such no-brainer things to do, just sincerely, it baffles me as to how there cannot be evidence regarding this. Can the developer -- Is there really no evidence or you just couldn't find it? I mean, this doesn't make any sense.

CO-CHAIR KNOWLTON: Before we go to the developers, let's stay -- I will go over to them, because I think Peter wants to answer your question. Peter?

DR. SCHMIDT: This measure actually has a major design flaw, in that if you counsel somebody and are not effective, then you can counsel them again; whereas, if you counsel somebody and they stop driving, they fall out of the measure. They fall into the exception criteria.

So if you continually counsel your panel and they continue to drive, you can get a perfect score on this.

CO-CHAIR KNOWLTON: Works for

smoking, too, doesn't it? Did you have another point, Ann? Okay. Fred?

DR. TOLIN: Looking at the prior measure and this measure, 2029 -- and I know that some comments were made in a separate -- in a Work Group discussion about this -- it is a little unclear to me why this was singled out as a risk factor when the other measure is more globally looking at a bunch of risks, and I was really curious as to why that might be the case.

CO-CHAIR KNOWLTON: I understand, and I certainly understand, A.M., the concern you stated about just sort of driving in general.

DR. TOLIN: I think we all feel, too, or logically understand where this is coming from, but I was just really curious s to why this was singled out as a separate item when it would otherwise have been inclusive in 2028.

CO-CHAIR KNOWLTON: I would just

be guessing, but it would seem to me that, to Salina's point, there's probably many, many, many more people driving than using guns or handling toxic chemicals or working as electricians in these circumstances. So there is much bigger and a lot more people that you are dealing with. So I think, in hindsight, that is why they split it out, but they included it in the other one. So I don't know why. It seems duplicative.

Anything else on the measure? Okay. Oh, I'm sorry, you are right. I forgot.

MS. TIERNEY: Your question about the evidence: I just say this to emphasize what I said earlier. We based our measures on the practice guideline. So we are limited to what the practice guidelines include and their summaries of the updates.

We do not do systematic evidence reviews, similar to other measures offered, and NQF's Task Force report, although they had

indicated a higher bar required for the evidence, they do specifically state that they don't expect developers to conduct primary systematic evidence reviews, but rather to report on those done by others.

So those are kind of our limitations. So the guideline with AAN are just on the guideline, probably many of you are aware in 2010, on driving with dementia, and that guideline has rich information about the evidence available, but they looked at specific questions we have been trying to answer like what tools might be useful for identifying patients at increased risk.

Our measure -- The evidence provided to support that in that response, to answer that question, doesn't necessarily address -- focused on counseling, mentioning alternatives to driving. It is a very patient centered measure, and some of the evidence that is supported by the AAN guideline, research questions which don't necessarily

directly link to our measure, and so our measure doesn't incorporate those.

If our measure was for assessment of whether the patient was at increased risk on driving, then we included them, but our Work Group felt that the measure was more appropriate as a counseling measure to increase visibility of the issue.

Just to your point, someone earlier was questioning kind of how this could be included in both measures, I can see that it does appear a bit redundant. With the safety concerns measure, that measure is intended to be very broad, and we contemplated leaving out driving, because it is covered by this measure, but we didn't want to necessarily send the wrong message, that driving isn't also a safety concern.

So since that measure was so comprehensive, we felt like we should include all of the various elements that are appropriately of concern, and the part of the

reason driving -- that we didn't just entirely include it all that one measure was because driving, unlike the other basic concerns, has a significant potential ramifications on the safety of others. So we felt like it warranted its own measure.

The little information that is available in the literature related to the gap in care does show that certain elements are not being consistently done in clinical practice.

CO-CHAIR KNOWLTON: Okay. Jane?

MS. SULLIVAN: Just to build on Salina's point, do we not have data from public safety about the percentage of people who have dementia or who are receiving care for certain things who have traffic accidents?

It would seem like -- I am just thinking about all these safety concerns. Do we not know something about the incidence in this population of safety issues, and could we not consider that kind of data?

DR. BARRETT: There is some of that data.

CO-CHAIR KNOWLTON: Go ahead, David.

CO-CHAIR TIRSCHWELL: I was just going to comment that there may be evidence out there. It is not our job to present it, and the problem is that it was not presented, by and large, in the applications.

I have to say, it goes back to the Psychiatric Association's guidelines that they are reviewing, which don't present the data in detail either, and they didn't write those guidelines as ammunition for NQF quality measures either, but it is all part of this environment where this is what it is being used for.

So I think the message has to go back that, as these measures are being developed, that is the kind of evidence that NQF is requiring, and so the care provider organizations that are involved in the care of

all these different types of patients with these different measures have to be able to provide what is needed to support their measures.

CO-CHAIR KNOWLTON: John?

DR. DUDA: Sorry, I am just curious. If you look at 1(c)(16) where it says that all patients and family should be informed that --

CO-CHAIR KNOWLTON: I can't hear you. Speak in the mic.

DR. DUDA: Sorry. If you look at 1(c)(16) from the APA guidelines, it says that all patients and family should be informed that even mild dementia increases risk of vehicular accidents, Category 1. There is a bunch of those. What does that Category 1 refer to, if it is not evidence based kind of delineation?

MS. TIERNEY: The Category 1 refers to a recommendation based on categorization here and judgment of those

patients, and it is included in your form what that refers to.

CO-CHAIR TIRSCHWELL: Something like they have great confidence, clinical confidence. Well, you know, it is a good question, and there are other grading systems that are out there that are much more specific about what it means in terms of trials and things like that. I think that is -- This conversation suggests that maybe that needs to be done in a little bit different fashion going forward.

MS. TIERNEY: If I could just add -- and I agree with your point -- just a little bit more about -- The guidelines themselves are included at 1(c)(10) of the document. They indicate that each rating considers the strength of the available evidence and is based on the best available data. When evidence is limited, the level of confidence is also incorporated.

I just wanted to add that.

CO-CHAIR TIRSCHWELL: Anyone else?

DR. BARRETT: Yes. I just had a follow-up to one of the comments made by the developer, again related to the difference between the measure being a counseling versus assessment measure.

If I can just clarify and ask the developer. It was the view of the Work Group that counseling would have a larger impact on public awareness? Is that correct?

MS. TIERNEY: Yes. So the Work Group felt like a counseling measure might do more to potentially impact the problem. You could involve caregivers in that counseling, like as many as 12, to kind of highlight the potential safety concern.

DR. BARRETT: The follow-up comment I would make then related to that is that, because of the requirements of the NQF process, that invokes another logical step that needs to be investigated. So then the effect of caregiver awareness needs to be

evaluated with respect to its impact on reduction of accidents and alterations in driving behavior.

Although a direct assessment measure of -- So a measure of assessment of driving competence, let's say, may be an easier measure to present to an organization like NQF.

CO-CHAIR KNOWLTON: Peter?

DR. SCHMIDT: I think that we are overly critical of counseling guidelines. Drug prescription is also counseling. You are telling the patient to go to the drugstore and to get the drug and to take it, and if we are not measuring compliance, then you are really talking about how well does -- You know, I personally am a highly noncompliant patient when I am -- you know, I don't fill prescriptions. I don't take them.

So if somebody gives me a prescription, they are counseling me to take it, and I often ignore them.

CO-CHAIR KNOWLTON: That is right. Therapy can help you.

DR. SCHMIDT: Yes. So counseling is counseling. You know, you can't just say that counseling is a terrible thing and that we don't know how effective it is, because I can tell you that drug prescriptions are not that effective with me.

CO-CHAIR KNOWLTON: We will have a session right after this. Gail?

DR. COONEY: I do think, you know, t his measure as opposed to some of the others we looked at, does have data regarding the incidence of the problem. In 1(a)(3) they talk about twofold increased risk of crashes, impact on driving that increases with dementia severity. So there is, actually, some data and some public health issues.

CO-CHAIR KNOWLTON: But that is impact data. That is not evidence. Not evidence, in the evidence that would deal with the reliability of the measure or the validity

of the measure.

DR. COONEY: Correct.

CO-CHAIR KNOWLTON: So it is an impact issue. We are going to actually get to that later, but it is an impact issue. Jolynn?

MS. SUKO: I think just another comment about counseling measures. I noticed that many of these are designed for physicians, and I vaguely recall from 15 years ago sometimes the counseling measures aren't always -- such as these, they are more effective when given by other caregivers such as social workers or nurses.

So I would just challenge -- and I don't know what the literature is, but when I look at this, I would ask, are we adding yet another burden to physicians for something that they need to do, when indeed it may not be the most effective use of their skill set?

MS. TIERNEY: If I could just add that the specifications for these measures do

include -- they are applicable to other care providers, including psychiatry, psychology and social workers.

CO-CHAIR KNOWLTON: Any other comments? Jane?

MS. SULLIVAN: Well, I am just Google Scholaring things. If you look for counseling, driving and Alzheimer's -- and I haven't read all these articles, but it seems to me that there's quite a few articles here that at least, just on my brief review, are suggestive of some interaction between counseling and a safety benefit.

So I think there is some literature here. It wasn't our job to find it. Googled Scholar counseling, driving and Alzheimer's.

CO-CHAIR KNOWLTON: But as David appropriately pointed out, that is not our task.

MS. SULLIVAN: I agree.

CO-CHAIR KNOWLTON: But I think

the same with these measures all along. One of the frustrations from where I sit is, speaking for myself, that these are clearly important things, and when we don't treat them, we don't get them to the level of an NQF measure, and it is a lost opportunity for dealing with something that, if it is that important, we should be doing.

So we got to put the rigor in to get it done. That is where I get frustrated with these, because it is not that I don't think -- A lot of people die in motor vehicle accidents, because they are not competent to drive, and every primary care doc knows it, and everybody treating people.

Every ER doc assesses for it, but we are not rising this to the right level is what is frustrating, because it does need to be risen to that level, but if it doesn't get to the level, it won't get done in an orderly way. That is the frustration. Mary?

MS. VAN DE KAMP: I think this is

-- If we looked going forward, this has been systemic of our industry in clinical practice. I think there is an assumption that clinical practice is based on evidence, and so, therefore, we haven't always judged the evidence, and I think that speaks to many of the outcome measures that we have struggled to try to defend or bring forward, is that just outcomes and evidence has -- We have done a lot of things, because we thought that they worked or were pretty sure that they worked and there's evidence. So I think it is definitely a part of our industry that has taken on a big clinical practice about the evidence, is the challenge to bring to the highest level.

CO-CHAIR KNOWLTON: Salina?

DR. WADDY: This is really a question, I think, for the NQF, and is there an opportunity for these two measures in particular for them to go back and reassess and to determine whether or not there is

appropriate evidence outside of essentially the guidelines. Is there that opportunity?

CO-CHAIR KNOWLTON: Helen?

DR. BURSTIN: That is actually what I was going to say when he called on you. So if you look at the why we have written out the Noes, there are two noes there, and one of them is intentional to sort of get at the issue of there isn't evidence, and you have been talking about this exception a lot today.

The other one is that there is insufficient evidence submitted. It may be out there, but it wasn't submitted. So one opportunity might be, as you vote today, you should feel free to invoke 3, and then we would ask the developer to bring that evidence forward, but then it is a judgment call, relying on the expertise of the people in this room.

CO-CHAIR KNOWLTON: We are voting on evidence. Let's vote.

MS. THEBERGE: Zero, Yes; five,

No, evidence does not meet guidance; 19 No, insufficient information submitted.

CO-CHAIR KNOWLTON: Okay. We are done with this, and we saved the best for last. Gail?

DR. COONEY: I love going last.

CO-CHAIR KNOWLTON: Hold on for a minute, Gail. Yes, Salina?

DR. WADDY: I was also wondering, in light of that, can you just tell us what the last vote was, the split between the two and three?

MS. THEBERGE: On 2028, it was one Yes, No 10, and then insufficient evidence submitted.

DR. BURSTIN: They are welcome to submit additional evidence, but then it will be up to you to decide if you want to reconsider it.

CO-CHAIR KNOWLTON: But we do that later when we see the evidence, not now.

DR. BURSTIN: Yes.

CO-CHAIR KNOWLTON: Gail?

DR. COONEY: This is measure 2030 which is also part of the AMA-PCPI project, and it looks at the percentage of patients, regardless of age, with a diagnosis of dementia whose caregivers were provided with education on dementia disease management and health behavior changes, and referred to additional resources for support within a 12-month period.

The denominator has the exclusions for medical reasons, including severe disease or no caregiver present, and the evidence is pretty much the same that has been presented for the earlier measures, for the structure process outcome.

They do have evidence that greater caregiver knowledge is associated with higher care quality, that intensive caregiver support resulted in improved patient outcomes, such as delayed nursing home placement, and that providing additional resources to caregivers

is important, is critically important.

CO-CHAIR KNOWLTON: Questions for Gail on the evidence? Gail, did the committee have a sense on the evidence? The sense was that it is not there yet, or where are you?

DR. COONEY: The committee's sense was pretty much the same as it was on the others, that the evidence was insufficient to support the measure.

CO-CHAIR KNOWLTON: I had one comment on this in that I was concerned about caregivers, not that -- I agree with your presentation, Gail, but I agree with the importance of getting some good measures here, because I think it is important.

Caregivers are very variable, and I think there is a disparity issue here with poorer patients not having access to necessarily good caregivers, and where that gets stratified, you can have -- I guess I am just making the editorial comment for the purpose of the transcript that we are not

doing a real good job with caregivers, and trying to figure out the wide variation there.

That wide variation is especially disparate based upon income level and race, and I hope s developers listen to this and get involved in this, they will pay attention to caregiver is an important variable, and we owe it some diligent sight, I guess is what I am saying. Ann?

DR. BARRETT: I think I am just echoing and emphasizing your comment, that studies reveal that many caregivers don't even understand that their love one has Alzheimer Disease or dementia, much less the particular activities that need to take place in order to optimize that person's quality of life or reduce their own burden.

So the public health need is very great. I agree with you.

CO-CHAIR KNOWLTON: David?

CO-CHAIR TIRSCHWELL: I just had a question as to -- Many patients with dementia

live in nursing homes, and so who are the caregivers then, and who is going to be rated on this? It doesn't seem like it would be the family that comes in once in a while. Is it the nursing home staff? Is it appropriate to apply to those patients?

CO-CHAIR KNOWLTON: Are you being responsive for that? A.M., go ahead.

DR. BARRETT: I am happy to be corrected by those with more experience with those patients, but in general, medical care and counseling don't end with skilled nursing placement, and the needs of caregivers continue even after the point of skilled nursing care.

In fact, some studies indicate caregivers have more needs during that period of time, because it is not clear to them what their responsibilities and interventions could be.

CO-CHAIR KNOWLTON: I don't know if NQF has seen other measures in caregivers,

but this is such an important area. I hope people pay attention to it. Risha.

DR. GIDWANI: I was a little bit more comfortable with this measure, really, because of the fact that, when the developers were citing the guidelines, they did mention that studies indicate that education and support for caregivers increases the likelihood that patients are adherent to treatment recommendations.

That, to me, seems like an important link that I would like to see explored a little bit further. If those studies do exist, I think that it would be really beneficial for us to be presented with information about their quality. Even if the guideline developers aren't showing that, if the -- I'm sorry.

If the developers of this measure are able to actually able to go to those studies themselves and give at least some sort of brief overview of what are the outcomes

that were measured, and was it a prospective or retrospective or randomized or observational study? Even that basic level of information would go a long way in helping us to better evaluate the body of evidence.

CO-CHAIR KNOWLTON: Anybody else? Okay, on the evidence.

MS. THEBERGE: Zero, Yes; six, No, evidence does not meet guidance; and 18 No, insufficient information submitted.

CO-CHAIR KNOWLTON: We now look for NQF member and public comment. Arnica, can you open up our phone line and see if there are questions for us?

OPERATOR: At this time, there are no questions.

CO-CHAIR KNOWLTON: Thank you. Yes? Robert?

DR. PLOVNICK: Hi. I am Rob Plovnick. I direct the Department of Quality Improvement in Psychiatric Services at the American Psychiatric Association. We were one

of the groups that worked on the development of these measures.

I just wanted to comment on a few things. First of all, with regard to guideline development, I just want to remind the group that there is a considerable time that it takes to develop guidelines and performance measures.

So there is a lag in guideline development and measure development, and these guidelines were actually developed several years ago, and we certainly have revised our guideline development process to be more aware of new IOM guidelines in terms of grading the strength of evidence and other factors that would tie into measure development.

So, hopefully, these types of conversations will be easier going forward. That being said, for many of the recommendations here, were we to have explicitly graded the evidence, it would have been weak.

There are not randomized controlled trials for assessment and for counseling. Dementia and Parkinson's Disease are degenerative conditions that impact significant segments of the population. This group is aware of that. This is really why this group was convened to assess measures in those areas.

For these type of conditions, the desired outcome is not resolution of the disorder, but optimizing quality of life, addressing new and emergent symptoms as they emerge, and then treating -- providing treatment that is compatible with patient and family wishes.

Gaps in care that prevent these types of outcomes are insufficient assessment of symptoms and management of them over time, and of counseling. It is unlikely that we are ever going to have strong evidence randomized controlled trials on these aspects of treatment.

I think that would apply to these disorders and other degenerative conditions. So I just want to note that, if it is desired to have meaningful measurement, performance measurement, for these conditions, that we might need to review this process in terms of how evidence is considered.

CO-CHAIR KNOWLTON: Any other comment here? Salina?

DR. WADDY: Even though we would like to have clinical trials, we haven't excluded other types of quality research, and what we are asking for is quality research to support the things that we are doing so that we can implement in a knowledgeable way across the U.S.

So even though a bar for high quality clinical trails is important, that is about the only thing that we are considering.

CO-CHAIR KNOWLTON: Risha. Peter, did you have a point? I will piggyback on what Salina said, that the issue is are we

reporting the same thing consistently and with enough rigor of what we are trying to measure.

I thought Peter was comment on his repeated comment earlier of randomized controlled trial, while a very high bar, is not the only bar. There are many other ways to have reasonable and appropriate research, and randomized controlled trials is just one of them. David?

DR. HACKNEY: I think a very high bar for evidence is important, because when you establish a standard, you fix practice into that standard that people believe they are obligated to do that; and if you haven't shown that not only is a reasonable thing, but it is the right thing, you may be actually, at least at the applied clinical level, cutting off innovation and forcing people to do things that no one knows whether it is really the optimal method.

CO-CHAIR KNOWLTON: Or we would still be doing bleeding and cupping. You mean

they stopped? Anything else? Yes, Mary?

MS. VAN DE KAMP: One thing. I am just wondering if there would be any information that comes from the PQR measurement that is already in place. Does that come back to help us then make evidence, because some physicians are already doing some of these measures, and we are tracking that through CMS.

To me, I don't know what we look at for that, but if we don't pass this, where does that information go? Does it come back? How does that work?

DR. BURSTIN: We have been trying to actually work with CMS and the developers to try to break that log jam and get more of that PQRS data flowing to the developers. It has been a challenge. But again, part of the issue is also participation is somewhat low in PQRS. So the numbers tend to be small.

It would still be helpful just in terms of getting even a validity check on the

rates of performance.

MS. TIERNEY: If I could just add, related to that question about PQRS. One of the challenges for us as developers is that that information that we get, even when we get it, only speaks to the gap.

So we already have some data from the medical literature related to the gap. That would provide additional information related to the gap, and maybe be more nationally representative, possibly, although it is a voluntary reporting program, but it still won't solve the evidence problem.

I don't know how that can be solved, to add to Robert's point, for a condition like this. I think that our Work Group that developed the measures tried to identify those areas that they thought physicians could improve upon, and that would lead to improved care.

They selected appropriately in many things dealing with assessment and

counseling, and I can appreciate many of the comments related to the evidence bar needing to be high, but then it almost seems that a condition like dementia will never have any NQF endorsed measures, because the evidence will not be there.

CO-CHAIR TIRSCHWELL: Just as a response to your "it would only inform the gap" statement, you know, CMS has all sorts of outcomes data, rehospitalizations, use of resources, costs, hospitalization, mortality; and if you have information about who is or isn't getting these measures done, it seems that observational studies are begging to be done.

it is not your role to perform those studies, but it certainly seems like there would be the ability to create some evidence that looks at outcomes related to these process measures.

DR. SCHMIDT: I think that one problem that we have is that PQRS has a

tremendous availability bias. So it is a voluntary submission. It is going to have -- As far as evidence, I don't see that I would review a paper written based on PQRS data as evidence for whatever was being measured in it. That is a challenge.

I think that one thing that we identified as a group is that a number of these measures that went -- A number of these assessment measures were really submitted backwards. Somebody should have said, once this is identified, evidence supports this therapy, and then that would give us information that could inform -- That would give that group information that they could use to write an observational paper on assessment.

So I think that we all anticipated with some of the dementia measures that not passing the assessment measure would bring down the intervention measure, but as the discussion -- I think they both went down, but

as the discussion unfolded, it was the intervention measure that got more support. I think that developers should think about that.

Then my third comment is I think that many of the developers would not have anticipated the negative position that the group has taken on counseling. They wouldn't necessarily have thought we need to include a study on the efficacy of counseling.

In 2004 I was reminded that I offered an anecdote about my own case, but there was a study done at Kaiser where they randomized people into people who were given a prescription for Zyprexa and people who were given a prescription for Zyprexa plus counseling.

The compliance -- The six-month compliance was 37 percent for the people who did not get counseling and 74 percent for the people who did get counseling. So there is evidence out there for efficacy of counseling,

and presumably in the future people should include that sort of thing in their submissions.

CO-CHAIR KNOWLTON: I think also there is the concern, though, of what is counseling. Is counseling -- Remember that the measure has to have some stability to it. So it is reliability, so that I can say that what is being done over here by William is the same as what is being done by Anna, that that is counseling.

So is it counseling to say, you know, you really should stop smoking? Is that counseling? It doesn't even get to that. Sometimes counseling is a check in a box.

Yet you can establish, as you know, Peter -- You can establish some standards for that. Go ahead. You can respond.

DR. SCHMIDT: If I can respond, your point is -- Jerry O'Connor from Dartmouth with a cystic fibrosis study has shown that

differences in counseling about cystic fibrosis and how at different centers it is very difficult -- that the great centers and the middle centers both do counseling. The great centers do it better. That was written up by Atul Gawande in his book "Better" as well as the papers by Jerry.

So you are absolutely right on that, but we also -- You know, I think we need to -- We can't be too prescriptive in how people practice medicine, and I think that people should look at the way that drug guidelines are written up.

So, for example, a very effective one has been aspirin for CABG. The way that that is written up, it doesn't say you have to tell your patient that they should take one aspirin every other day or a baby aspirin every day. Tell your patient to take aspirin, and the assumption is they are going to be told to take it the right way.

So we need to think about, when

there is evidence, that it can be done wrong like there was with the depression screening in Parkinson's. I take that as we should be prescriptive in those situations, but when there is no evidence that, when this is done, it is done wrong, is that our role to assess that?

CO-CHAIR KNOWLTON: I can't resist responding to you. I think that our job is to be prescriptive, not to be prescriptive in medicine -- I agree with you. I think our job is to be prescriptive when we say we are going to measure performance and publicly report it. We have an obligation to be prescriptive then. We have an obligation that you should know that what I am measuring you on is identical to what I am measuring Gail on. But in terms of clinical practice, I don't want to interfere with that. But if I am going to measure it and I am going to say you got to do it, and I am going to report whether you do or not, with all the incentives or disincentives

that could be tied to that, that is where I think the rigor is.

There are all kinds of things in clinical practice that people make judgments on that we shouldn't interfere with. I agree with that. Salina.

DR. WADDY: I just wanted to go back to David Tirschwell's previous point regarding the use of CMS and regarding the use of actually big data.

One thing that is currently going on is CMS actually working more closely with NIH in order to see whether or not there are questions, potentially as this, that can be answered and made more available to outside investigators. That may be a way of getting the information that you mentioned.

The second thing is also to engage other groups, such as Kaiser Permanente and other large practice organizations, in order to put in place very simple interventions that can then rapidly be studied.

Those are very simple ways you can engage those types of organizations, but it is not our job to develop the research projects. It is to assess whether or not that has reached the level of evidence that can be more broadly distributed.

CO-CHAIR KNOWLTON: Ramon?

DR. R. BAUTISTA: I would tell the developer potentially to study number 209, which we actually passed yesterday. It is an example of a study that actually did not use RCT and got passed.

In fact, their first statement says evidence does not exist, blah-blah-blah. IN other words, they did not have any direct evidence for their thesis here; yet had a lot of what you call circumstantial studies, direct studies looking at different aspects of the same problem, which again is a compelling argument to pass this.

So it is a good example to study that is not directly RCT. It has very good

evidence, I think circumstantial evidence for their study. So it is a good study to actually look at and model for future reference.

CO-CHAIR KNOWLTON: We are going to move on now to measure gaps. That is, through our discussions today, have we elicited gaps that we think should be considered for future consideration?

MS. JOHNSON: I think, along with that -- and you guys have already started doing that, but maybe go ahead and put your measure developer hat on. These measures on dementia went down, most of them did. What wouldn't have gone down in your mind? What would be a good start? Let's give the developer some concrete ideas, not just dementia, but maybe we can start with dementia, and then we will go through and segue into the other ones.

CO-CHAIR KNOWLTON: I was looking at you, Anna, because I thought you would

immediately have a suggestion. Go ahead.

DR. BARRETT: I think that previously I made the comment about assessment of driving in Alzheimer Disease, but that is one of a number of functional interventions or assessment measures, process measures, that could be evaluated in Alzheimer Disease and dementia.

Although rehabilitation is oftentimes not thought to apply to progressive disorders, yesterday we acknowledged the importance of rehabilitation in Parkinson Disease, and in dementia there are a number of different interventions from traditional rehabilitative specialties that have been shown to improve function.

So actual assessment and referral for treatment and intervention may be appropriate gaps.

CO-CHAIR KNOWLTON: Peter.

DR. SCHMIDT: The driving one seems like one where you could create an

outcome measure around that, and I think an outcome measure would be much more powerful. I am not aware of any evidence that people with a diagnosis of dementia should be driving. So, actually, getting people off the road would constitute an outcome.

Maybe I am not aware of all of the evidence around that, but I think that that seems like something that is sort of a no-brainer.

CO-CHAIR KNOWLTON: Jane.

MS. SULLIVAN: Maybe this is a no-brainer, too, but if you look -- If the developers look at, instead of did the measure pass or did the measure go down, but when it went down, there were several cases where people seemed to feel that the evidence wasn't there, that there was evidence but it wasn't sufficiently cited.

So I would hope that developers would look at those measures in particular, and say, you know, people around the table

felt like there was some evidence, but it wasn't presented to the committee, and it wasn't the job of the committee to develop -- or to find that evidence.

CO-CHAIR KNOWLTON: John.

DR. DUDA: I think we have all remembered discussions where, if measures were designed to be more specific with not such a big umbrella covering the whole -- you know, the neuropsychiatric encyclopedia. If it was just depression, it probably would have had a better chance of passing.

CO-CHAIR KNOWLTON: Dan.

DR. LABOVITZ: I don't have any good ideas. So I am going to just offer a comment. I think that what we are seeing here is an ever evolving and improving general process, just within the medical community.

We are now establishing quality measures using a very strict process. It was only fairly recently we started coming up with consensus guidelines. Now our guidelines are

getting better.

perhaps one of the next evolutions needs to be having the various developers, the American Academy of Neurology being one I am most familiar with, because I am neurologist and I belong, looking to provide funding to show that an intervention makes a difference, and then using that to drive quality.

This group is the last stop, and is being pressed to offer up suggestions for measures. We need data.

CO-CHAIR KNOWLTON: Gail.

DR. COONEY: This is a little bit what John was referring to, and I doubt that there is evidence for this, but looking specifically at advance directives being written for dementia patients early in the course of their illness.

In my work, I too often see patients who missed the opportunity to have that advance directive discussion before they lose their cognitive abilities, and I think

that that would be very valuable.

CO-CHAIR KNOWLTON: David?

DR. HACKNEY: I want to second Michael's point from earlier in the discussion -- now I can't remember which measure it was -- saying that every new thing that you mandate happens during a visit is either extending the length of that visit in order to incorporate it or it is crowding out something else.

So I think, once you have declared that -- If you are going to declare that something is so important that this has to be done, then you should have evidence not only that it is useful in its use, the desired purpose, but some idea of what the magnitude of that impact is, because you could spend literally all day with each patient if you fill out every validated measure tool that there exists that might be relevant, and particularly when you are talking about elderly patients with multiple problems.

There could be an infinite number.

Physicians make the decision of what has to be done and what doesn't, but if you create a long list of mandates, it better be true that every one of those is important enough that it has to be done the way we are specifying.

CO-CHAIR TIRSCHWELL: Terry.

DR. RICHMOND: This might be going too far afield, but I am going back to the safety concerns with dementia and also the driving issues.

I think in health care we tend to be a rather incestuous little group, and I think that there are some really solid data out there, if we would expand our horizons. So NHTSA, National Highway Traffic Safety Administration, has a wonderful accident analysis reporting system with solid data. It is a regulatory agency. They will have data on causes of fatal accidents and a random sampling of nonfatal accidents.

The same thing with one of the measures that is a counseling measure, for

example, was referred to counseling about guns in the home of demented patients. CDC has a national violence reporting death system which covers data from about 26 states where there really are data out there.

So we may need to like take off our blinders and look at who else should we be connecting with to get the data to look at things that, as health care providers, we can intervene to improve outcomes for.

CO-CHAIR KNOWLTON: Just piggybacking on that, DOT has an awful lot of information on the driving issues that are at question, too. Risha?

DR. GIDWANI: To echo what David said just now and what Michael said earlier, I think that the opportunity costs of adhering to these measures is something that it would be great to have some data on.

So, for example, when I was reviewing the counseling regarding safety concerns for the dementia measure, and it said

that the physician or the provider should be having a discussion with the patient on a number of different bullet points, I really do wonder about the time that that takes.

I think that, for the purposes of NQF evaluation, I am not sure what NQF feels about this, but I would love to see some information about how much time this actually takes. I think that could be easily done with a pilot study to say this conversation took on average six minutes across a sample of 20 patients.

I think that one of the things that we need to be concerned about is the fact that, when we are focusing attention on certain measures, that the attention that will then be focused on providers on meeting these measures is not inadvertently causing quality of care to reduce on other conditions that don't have those measures associated with them.

CO-CHAIR KNOWLTON: I am going to

call on Jane. Gwen, weren't you going to say something, or not? Did somebody already cover it? I just made you wait too long. You gave up.

DR. BUHR: Yes. I was going to respond to Peter about -- He was saying an outcome measure about driving and dementia. I don't know that it is that black and white, because somebody with mild dementia with no impairment in their executive function may be able to drive; whereas, somebody else wouldn't. So it may be more complicated.

CO-CHAIR KNOWLTON: Jane?

MS. SULLIVAN: This is to build on Risha's point, which build on Michael's point, which was related to David's point, and it is the whole -- the time issue, but I think I would like to broaden that to not only who has the time but who is the appropriate provider to be doing some of these things.

The focus of much of this has been the physician. I would argue that a

functional assessment is probably as appropriate or more appropriately done by a physical therapist or an occupational therapist.

So to broaden the definitions of these things to look at who is in the best position to do it and who appropriately has the time. I think that goes to burden of care, but appropriate provision of care.

CO-CHAIR KNOWLTON: I want to piggyback on that question for NQF, raise with Helen or with Karen or with Suzanne.

I remember this issue being a big issue in the first round stroke, which I co-chaired as well, and it was particularly on dysphasia screening and who was the one that did the dysphasia screening. Mary is nodding, because Jane and Mary remember the debate.

I thought the argument at the time was that we didn't have to be prescriptive, because the standard is open-ended. It doesn't necessarily justify the physicians.

It applied to nurses, but this is an issue carousel.

It keeps coming back, you know, where people say we may think it is from Mount Olympus NQF. We may think that this is understood in the field, but it is not. So some data gatherer is saying, if a physician with an MD or a DO doesn't write down something in a chart, if an advanced practice nurse does, it doesn't count, or if a PT does or an OT or a speech and hearing person does, it doesn't count.

I just wanted to piggyback on Jane's comment, because I remember this issue. We spent like a day on this issue. Mary?

MS. VAN E KAMP: The developer left for the dementia piece. I wanted to say publicly, and I think it is important, that the investment and the time to bring these forward is really recognized.

I think that, while it may not have all the pieces to it that we need to have

to get to a high standard, I think it is important that the discouragement from that may be work and not driven to the next level is not preventing continuation; because there is a lot of work that I know people put in this to bring it forward, and it is the first step in health care that we start to hold each other accountable and determine what we should spend our time on and how to do that.

I know that in the first phase there was -- I saw some disappointment, and I saw almost a stop in their continuation of bringing that forward on the ASHA NOMS. I think that is something we need to go back and talk about, but I think it is important that this isn't seen as a discouragement, although I can appreciate that it is, but that really a next step. What do we need to do to really grow this and to hear what is needed to have that done, because I do think, as I watch faces in the room after you have worked so hard an you feel that work just sort of fall

down -- and again I think we have said it before. It is not that it is not important. It is that we have to continue as an industry to raise the bar, but I applaud the groups that have brought things forward, because there are many people in health care that have been hesitant to bring forward some of the measurements for fear of being judged by them.

CO-CHAIR KNOWLTON: Michael.

DR. KAPLITT: My thought is more to NQF than to new ideas. The Neuro Committee is a new committee for you guys. Right? Last time was our first time meeting, and it recognized, obviously, the emergence and the importance and the increasing number of neuro guidelines, but neuro was different than a lot of other areas of medicine and can have a lot of these vagaries that are not as clear when you are dealing with very concrete data points like blood pressure and other things.

I think that a lot of the problem -- It is clear that there has been enormous

sympathy for a lot of these things among most of us, because we all know that these are important areas in general, but the struggle has obviously been that essentially what we have been asked to do for many of these or most of these things is to largely rubber stamp guidelines that were made -- organizational guidelines without really any, much additional information.

My sense from the comments of the various developers, not one in specific because it seemed like a common theme, was not entirely a full understanding of what the NQF process or needs are, because repeated statements like, well, you are never going to do a randomized controlled trial on this, so this is the best we can do, to me seems dramatically divorced from what our process really is here.

So if they go home with that message that, well, you know, all they want is randomized controlled trials and that's it,

that obviously is a bad message. But we can sit here and say that all we want today, but it seems to me like maybe there needs to be more of an engagement in advance of this process between the developers and NQF to understand the process and what evidence means and what the purpose is here and what we are all trying to do, because this isn't -- I don't think anybody here is playing a game of gotcha like, well, you know -- because I have sat on a bunch of NIH grant reviews where that is what happens, where people just look for like some random line -- you know, you didn't do that; that's it, you are done.

I don't think that is what has happened here today. I think there has been great struggling, and that is why we have spent a lot more time than I had thought we were going to spend yesterday morning with my initial statement, but it was well spent, because we are all struggling with this, because we all live out there in this world,

and we all want to make this better for these patients.

On the other hand, we also live in the world where we and our colleagues are increasingly under the gun, and that is only going to get worse. So we have both sides of that in our minds when we are doing this, and I think that, if the developer is engaged NQF a little more, particularly on the evidence side, and what we all want in advance, then that might help enormously.

DR. BURSTIN: Those are great comments, Michael, and we have actually worked really closely with developers to try to make this really clear, and we actually have regular monthly measure webinars. We actually have an in-person measure developer meeting in November.

I think there is actually, truly a bit of a disconnect of the fact that -- Again, I think Daniel raised this point as well about the supply chain. I think, to a certain

extent, the expectation that a developer is going to do a systematic review is not likely, and if the guideline developer hasn't done it, it does put them at a disadvantage.

So this is something I think we need to work out, although I will point out, though, very interestingly, every committee, by the way, thinks that their area is not as clear, and there is a lot of -- So GI, urology, we run through all of these. My patients are sicker. The mantra continues, but I hear you.

I think the other thing that has been really hard for us is that these are gap areas for us. We have almost no measures of dementia or epilepsy. So I think it is very heavy hearted for us as well. We want to be able to bring forward something that we feel like would really help move the field forward without - you know, you can only measure what you -- You can only improve what you can measure, and we have nothing.

So any ideas about what you think the developers could do to improve and try to move some of this forward -- and Parkinson's as well, obviously.

CO-CHAIR KNOWLTON: Salina.

DR. WADDY: I really want to -- This is really for the NQF, but I would like to go back to my statement from yesterday, which is there is a lot of valuable information, even if some things went down in flames, but there is a lot of valuable information in the discussion.

If there are certain elements or certain measures or areas where there are gaps that the NQF really thinks needs to be further investigated, then potentially sitting down with the agencies that fund research as well as bringing in other stakeholders just to discuss this is identified as a major gap area, and then review this valuable information that you have already gotten together, could be tremendously helpful.

DR. BURSTIN: We have had similar discussions with PCORI as well. So I think there is interest in seeing that. We used to actually have a section of the report, I think, on research recommendations. Maybe it is time to kind of loop back to that, as long as we are clearly having a tough time getting some of this through.

DR. WADDY: Right, but I am involved with clinical research. For example, we do bring in outside groups if there was an interest to talk to the entire group.

DR. DUDA: It seems to me that a lot of the problems that were raised in the small group conference calls about these measures came to us for this meeting, and perhaps if there is a longer delay in between when they get that feedback back from the small groups until this meeting, they could have more chance to answer those queries or respond to those criticisms.

CO-CHAIR KNOWLTON: Jocelyn.

DR. J. BAUTISTA: Getting back to the original question of what might have passed, in terms of epilepsy we have said multiple times that NQF prefers measures that are close to the outcome. So we reviewed a couple of measures that asked for documentation of seizure type and seizure frequency.

So a measure -- and I think the developers actually mentioned this yesterday, but a measure that said for patients who are not seizure free, what percent are referred to an epilepsy specialist or what percent are referred for surgical evaluation, something that does more than the assessment, but actually acts upon the assessment.

CO-CHAIR KNOWLTON: Peter, nothing additional? Okay. Anybody else? Then I suggest that we break for lunch, and then we will come back for our additional discussion topic Phase I follow-up, which is on the Yale/CMS stroke measure.

DR. BURSTIN: The disparities-sensitive measures would apply to those that went through. So since they were few, we could see if there is any specific interest in any of the ones that did get through.

MS. JOHNSON: I think, just a reminder, too. My understanding is that the ones that went through were good: the two dementia measures about diagnosis in nursing facilities and the counseling for women of childbearing potential with epilepsy, and then the stenosis measurement.

So the first two of those, you have already told us that those are disparities-sensitive. So I think our discussion -- That would have been our discussion. I guess the other question then would be for the carotid imaging studies.

CO-CHAIR KNOWLTON: Let's eat.

(Whereupon, the above-entitled matter went off the record at 12:07 p.m. and resumed at 12:51 p.m.)

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

(12:51 p.m.)

MS. JOHNSON: Everybody, let's go ahead and get started back again, in the interest of time and, hopefully, we can finish up our discussion this afternoon.

To start the next section of our meeting today, we are going to go back and revisit the stroke readmission measure from Phase I. So to start us out in our thinking about that, I am going to turn it over to Helen.

DR. BURSTIN: Hi, everybody, again. The last task before you get to leave us, we want to just take this opportunity to thank you for, obviously, all the hard work you have done the last couple of days, but also to explain why we are revisiting this measure, because I know some question has come up.

So just to recap, just historically, this measure was discussed by

you the last time you were in this room and in this meeting, the first phase of the project, and the measure was approved. The readmission measure was approved.

I will also point out just to remind you, the mortality measure that we talked about, the 30-day mortality measure, has been withdrawn by CMS, and they are now investigating other approaches potentially to get more clinical data like the NIH severity scale, part of it, moving forward.

So one of the issues is, when we had that follow-up call with you, you were voting on the updated measure that had included the expanded planned readmission algorithm, as well as the expanded risk adjustment age categories for the mortality measure.

So we had you revote on that measure at the time and, if you recall, much of the discussion we had on that conference call was really focused on the fact that the

Fonarow paper had come out, and a lot of the discussion was heavily focused on the relationship of the NIH severity scale to the outcome measures, primarily focused on mortality.

Because the measures had substantive changes made to them, we are required to put them back out for public comment, since people didn't have an opportunity to see those updated measures. And since we were sending out the mortality measure, we elected -- even though the vote was very close, it was 10 to 12 -- to just put them both out for comment, particularly since it had already passed initially before that conference call.

So at this point of the discussion, we are going to have an opportunity to review the public comments that were submitted. We will then have Yale and CMS have an opportunity to respond to any of those comments, and answer any of your

questions.

One of the things I just have to acknowledge is at times it is very difficult for us to know exactly what constitutes consensus. So votes that are that close and so nearly split are ones we tend to err on the side of getting more information, putting it out for comment, getting as much information as we can to bring it back to the committees.

Certainly, if that was the final vote, it would not go out for voting. It would stop there. So voting requires truly this majority rules, and it just goes out. So we actually have a Consensus Task Force now that has been convened by our Board of Directors to help us really kind of more crisply define what we mean by consensus in terms of these votes.

So at the conclusion of this discussion of the comments and specifically, again, trying to focus at this point only on readmissions, and so much of our discussion on

that conference call was so heavily oriented to the mortality measure and the NIH severity scale, we are going to focus this one today just on readmissions.

I will point out, as you saw in the AHA comments as well as communication that I got after presenting indirectly at the Brain Attack Coalition recently that there does not appear to be significant evidence that the concerns about the NIH severity scale apply to readmission, as best we know at this point.

So we are just going to focus on readmissions. At the conclusion of this, you will again have the opportunity, as we always do for committees whenever new information comes forward, to say is there anything as a result of this discussion that would lead you to want to revote again on a readmission measure.

So that will be your discussion point today, and at that point I will turn it back over to -- who is doing this part?

CO-CHAIR TIRSCHWELL: I am. I have one question. I guess I wasn't sure if the folks that had submitted the comments about the measure and might be calling in, and we are starting early. Is that an issue?

DR. BURSTIN: We have got most of the Yale and CMS folks in the room, and I know Kate Goodrich was going to be calling in. Kate, are you on the phone? Arnica, are you with us?

CO-CHAIR TIRSCHWELL: We are just wondering if there is anybody connected.

OPERATOR: Yes. There are several speakers on.

DR. BURSTIN: Okay, could you please see if Kate Goodrich is there and, if so, put her on the speaker line for us.

OPERATOR: She is not online at this time.

DR. BURSTIN: Okay. Please let us know when she is and, if not, we will defer to Lein Han from CMS as we need to. But we are

okay proceeding.

CO-CHAIR TIRSCHWELL: I guess my question was, was the Heart and Stroke Association planning on calling in?

DR. BURSTIN: Not that I am aware of.

CO-CHAIR TIRSCHWELL: Because they submitted a lot of --

DR. BURSTIN: Right. We have their comments.

CO-CHAIR TIRSCHWELL: Okay. I guess I am worried that, if the schedule had gone out that we were going to start at 1:30 and people were going to call in, they don't necessarily have to make a reservation, do they? So those people might miss their opportunity to participate.

DR. BURSTIN: I suspect we will still be talking at 1:30. So if they called in at the end, we will make sure we do public comment. How about that?

CO-CHAIR TIRSCHWELL: I just

wanted to make that comment. Helen, you said that we can -- this is a different process than any we have done before. We have a chance to review the comments, and I know the Yale group has a presentation to make.

Are you suggesting we literally walk through the comments or just ask people if they have particular things?

DR. BURSTIN: Do you want to briefly walk through briefly what we saw in terms of comments or have people had a chance to look at it, and people just want to make comments? I don't think we need to do a point by point on comments.

CO-CHAIR TIRSCHWELL: Yes, go ahead.

DR. DRYE: First, I just wanted to confirm if our Yale team -- I am Elizabeth Drye from Yale, and my colleague, Susannah Bernheim who led the measure development for both mortality and readmissions should be on the line. I just wanted to confirm.

DR. BERNHEIM: Yes, we are here, Elizabeth.

DR. DRYE: Great. We just wanted to do what makes the most sense to you. We didn't prepare slides on the comments, but we could -- Susannah could walk through and summarize. I just prepared slides to highlight a couple of points about the planned readmissions and also the medical record validation of the readmission measure, just actually five slides, very brief.

CO-CHAIR TIRSCHWELL: I will make a suggestion, and then people can suggest alternatives, if they like. I suggest we go ahead and let you go through the slides, and then open it up to questions related to that or any other topic that people had questions on in the question and answer document, if that is okay with you all. Okay. You guys want to go ahead and do your slide show?

DR. DRYE: Hi. thanks. We just decided to keep it brief. We know you have

had several different discussions at different stages about the measure. So I just wanted to highlight particular points about the readmission measure since, as Helen mentioned, a lot of the discussion has been focused on mortality.

The key change since we initially submitted the measure was to update the planned readmission algorithm to be more expansive, to identify more readmissions as planned, and it is a shift in how we are doing readmissions measures generally.

When we put the measure together, the team of experts, including neurologists and others expert in stroke, identified some readmissions as planned, that would be typically planned following an admission for stroke as related follow-on care.

So it was a fairly narrow definition and included things like carotid endarterectomy or intercranial sensing, and it was less than a percent of admissions that

were followed by planned readmission.

As we have continued in the measurement community to work on readmission measures, we wanted to identify a broader set of readmissions that were planned and, I think, as some of you know, we built an approach to doing that in claims data for the hospital-wide readmission measure.

This broader approach is seeking to identify as planned not only readmissions that are related to stroke and that are follow-on care, but just planned readmissions that occur in this particular population, and Medicare patients is the focus at our discussion, because they have a fairly high number of planned readmissions for unrelated things like cholecystectomy, for example, that might occur from 30 days of discharge from a minor stroke or other readmissions for other conditions like pneumonia, and we do not want to count those in a measure looking for a quality signal.

So what we did was we built an algorithm, and we prepared a report which I know was distributed to you, but it is complicated and lengthy. I just wanted to summarize really briefly.

We defined planned readmissions as readmissions that were for non-acute reasons. It couldn't be for an infection or a second stroke or a heart attack or any emergent reason, and that had a scheduled -- a procedure that we would call a typically scheduled procedure.

We never want to call planned admissions that are for acute illnesses or complications of care, and there were some kinds of admissions that could occur within 30 days of discharge that we have heard quite a bit about from people we collaborate with, from public comment, like rehabilitation, admissions for cancer chemotherapy, transplants, that really are planned and shouldn't be counted in this type of a

measure. So those, we don't count as planned.

So we have a list that we distributed earlier to potentially planned procedures and a list of acute conditions, and together those allow us to put readmissions in the planned or unplanned category.

When we applied it to the stroke measure, we had an expansion of the number of the percent of admissions followed by a planned readmission. So in the originally submitted measure, the readmission rate was 14.8 percent, the readmissions we were counting, and it dropped when we expanded the number of readmissions we are counting as planned to 14.3 percent.

That percentage of readmissions that were followed by what we are now calling a planned readmission within 30 days of discharge went from 0.6 percent -- that was the admissions following stroke that were closely related to the stroke care that essentially would be follow-on care for a

stroke admission -- and now we have a more expanded definition, and I will show you what it captures in a second.

When you apply in this cohort of patients, 1.1 percent of patients who are admitted come back with what we are now calling a planned readmission, and we are not counting it in the measure.

If you look, this is a little small, and I apologize. I try really hard not to put small words on slides. So let ;me just read it for you. But the most common thing for a stroke patient, the most common procedure for which they were admitted that we are counting as a planned readmission was endarterectomy, which was what we would expect.

Also, quite a few patients -- this is out of a cohort of 169,000 patients. There were about 800 admissions for endarterectomy, about close to 200 for diagnostic cardiac cath, and 180 for rehabilitation, 174 for

cardiac device related procedures, removal/revision of a defibrillator or pacemaker. It goes down from there.

You will see in the bottom of the list -- if you can't see, again I apologize; I will read it for you -- that there are some planned readmissions here for what I think about as essentially care that -- this is Medicare 65 and older patients -- care that these patients come -- they are happening subsequent to an admission for stroke in the 30-day window from discharge, and probably very or completely unrelated to that, and they just needed care. We don't want to discourage it. We don't want to count it in a readmission measure.

They include procedures like a colorectal resection, presumably for colon cancer. That is the most common diagnosis we saw with that procedure in this cohort, or a cholecystectomy, etcetera.

So I want to just pause there,

because we are using, again, a list of potentially planned procedures that we developed in consultation with specialists across the whole spectrum of providers, and we are using a list of acute diagnoses. If you have a potentially planned procedure but not an acute diagnosis, we will call you planned. That is an algorithm that isn't that easy to follow in two minutes. So let me stop and see if people have questions.

CO-CHAIR TIRSCHWELL: I have got a question. So are these additional planned admission procedures -- is the only change that was made to the measure and, if so, judging by the percentages you just gave us, to me, it seems like it is a small change that is probably not going to affect much of what the measure does.

DR. DRYE: I don't think it fundamentally changes the measure, if that is what you are saying. I think it improves the measure.

CO-CHAIR TIRSCHWELL: Can you quantify the improvement? I mean, it is half a percentage.

DR. DRYE: Yes. I can give you a little more information that might be helpful.

CO-CHAIR TIRSCHWELL: Is this the only change that was made, though? That was my first question.

DR. DRYE: We also specified the measure for all payer population, but that didn't change the measure. That was just additional testing in a California all payer dataset.

Then I just wanted to present one last slide, which is about the validation of the measure, irrespective of this change. But in terms of the effect, it is not conceptual. I think that we really don't want in a readmission measure to be capturing planned readmissions, and we were looking for a way to do that better.

We took work for a hospital-wide

readmission measure that looks broadly across the entire hospital, and allowed us to -- you know, in that context, we were able to develop this algorithm. It went through several rounds of public comment. Actually, it went through public comment again in the context of this process, and so it is just an improvement to better capture the underlying quality signal that we are trying to capture.

I think it makes it more fair, and will more fairly characterize hospitals. There is a small shift in how hospitals rank when you apply this, because they vary. You know, they look a little different when you count planned readmissions this way.

Does that answer your question?

CO-CHAIR TIRSCHWELL: Yes. Go ahead and finish your presentation, or did anybody else have any questions about that first part? Go ahead.

DR. DRYE: Okay. I am just going to highlight quickly the validation that we

did of this measure of the use of claims for risk adjustment for this measure.

It parallels what we did in the stroke mortality measure. I wanted to contrast, because the results are quite different than they were for the stroke measure.

We used the National Stroke Project medical record data, which contains a severity scale. It is correlated with NIH. Of course, it is not the NIH. It us. And we matched a set of patients, and we estimated -- Actually, we did it at the state level, not the hospital level, just given the number of patients that we had, and we estimated risk standardized rates of readmission.

When you use the medical record data we had for risk adjustment, and you estimate rates and then you use the claims data and you estimate the rates on the same patients, you basically get almost the exact same rates. The correlation coefficient is

.99, which is a lot higher than it was for the mortality measure.

I think that this is as expected, given the findings we had going into the study, which is we didn't expect stroke severity to really be a strong predictor of readmission. In our lit review, it was not identified as a predictor of readmission, and I know in the public comment, the American Hospital Association highlighted that they had found the same thing.

So my understanding of the primary concern about the mortality measure was the adequacy of the risk adjustment and the need for a stronger signal of stroke severity in the risk adjustment. I just wanted to contrast what we found here and what the literature says underlying that.

CO-CHAIR TIRSCHWELL: We are going to do questions one slide at a time. Is that what I should interpret your pause as? Oh, you are done?

DR. DRYE: Yes, I am done. Sorry. I am being too informal.

CO-CHAIR TIRSCHWELL: You had flashed some other slides, but maybe those --

DR. DRYE: No. They are just showing you -- I can show you real quick, if you want. Basically, they are just showing the distribution of the rates. As you, I think, intuited, they don't really change -- These are rate distributions before and after we extended the plan readmissions, and this is just a slide.

Again, if you look at the rates estimated with the original plan readmission algorithm and with the new one, you see that there are small -- and you subtract the hospital standardized rates with the old algorithm and the new one, you see that the rates are changing a little bit for each hospital, which means the order of the ranking will change, but that is what we expect. We think we have a more accurate, better measure

that is going to characterize hospitals more fairly.

CO-CHAIR TIRSCHWELL: So going back to your correlation curve slide, those two ways that you are risk adjusting, and then, I guess, what is on the vertical and horizontal axes is the risk adjusted readmission rate. Is that what that is?

DR. DRYE: The risk standardized rate produced by the claims based measure is the X axis, and the risk standardized rate produced by the National Stroke Project medical record data measure -- the risk adjustment was done with that chart extracted data -- is on the Y axis.

CO-CHAIR TIRSCHWELL: You said you did this at the state level, not the hospital level.

DR. DRYE: Right.

CO-CHAIR TIRSCHWELL: But you are not -- You think it would play out just as well, if you had the data, to do it at the

hospital level?

DR. DRYE: I do, and we have done it for other measures at either the hospital or state level, depending -- What we really want is a national representative sample, and those are expensive studies, and it is hard to get enough volume on those to do it at the hospital level, but given these results plus what is in the literature plus what we are finding with the American Hospital Association and we are finding in our work, we don't expect -- This is what we would expect.

CO-CHAIR TIRSCHWELL: I think, if your point here is that again showing that the NIH stroke scale or a stroke severity scale doesn't make a big difference in your ability to risk adjust for readmissions, I am willing, certainly, to admit that it seems like that is fairly well demonstrated at this point.

I guess my question is: Because the statistics in all these predictive models are so low -- I am sure it is true for the

medical record one as well as it is for the claims based one -- I am guessing that the correlation of these readmission rates with the totally unadjusted is over .9 or maybe you have that actual number; because if these adjusted rates are really no different than the unadjusted, then have we really done anything, and since we don't know the factors that are affecting readmission, are we really going to be able to treat hospitals that are somehow disadvantaged equally, or will they be rated as doing more poorly?

DR. DRYE: I'll try to touch both those pieces of the question. On the second part, will hospitals be rated poorly, we have looked at the measure, and this is in our initial application and it is true for all of our readmission measures: When we identify hospitals with a lot of low socioeconomic status patients, for example, and very few, there is a wide distribution of performance on these measures, including this one.

So there are some very well performing hospitals with a lot of low SES patients, and there are some not so well performing, and the same is true among those who have a more affluent population.

So when we see a range of performance, then we are accounting for the way that this measure is discriminating quality, and they are not those hospitals that are safety net hospitals or have more poor patients, more minority patients. They don't routinely look worse on the measure.

On average, if you look at the medians, they do slightly worse, but they have a broad range of performance. So that is why -- and I know this committee has already had this discussion -- we think we are not disadvantaging those hospitals.

To your other point, if you just ran -- If you had no risk adjustment and you just ran it -- I don't think we have done that. It is a great thought, but I want to

think for a minute about the purpose of risk adjustment, which is to level the playing field across hospitals that take different types of patients.

So what we are trying to do with the risk adjustment is not predict readmission really accurately. As you know, we are just trying to be fair, and hospitals that get sicker patients, who have a higher risk innately of readmission, we are trying to adjust for that.

So we are not a priori saying that we need a high C statistic in this measure, and we know, whether we use chart models or we use claims models for readmission, patient factors are not the whole story in predicting readmission, as you point out.

If you saw that they did very little, I am not sure that would change a lot. It would be interesting.

CO-CHAIR TIRSCHWELL: I guess, just as one response, and then, Risha, I would

be interested in your comments, even though the safety net hospitals are the ones with more low socioeconomic status have a range of performances, your comment that on average they perform worse concerns me; and if we were doing a good job, theoretically, of risk adjusting, then I guess I would hope that that wouldn't be the case.

Again, I think there is a risk of these hospital, on average, being rated as more poorly performing, and if there are then penalties associated with this, then it would be specifically the most vulnerable hospitals and patients that would be potentially financially disincentivized to improve their care. That whole scenario -- it is very theoretical, but it seems quite worrisome.

DR. DRYE: I think that is a valid question and concern, and the way that we think about it in the context of public reporting is, you know, I don't know what the truth is right now about those safety net

hospitals that aren't performing well or those better off, wealthier hospitals that also aren't performing well on the measure.

What we see, after adjusting for risk, there is this range of performance, and at this stage where there has been no public reporting, the goal is really to illuminate those differences.

As you know, it is an NCAP guideline not to adjust those things away, because if we adjust them away, we can't see them. I think that is the first goal. It would be great if -- you know, in an ideal world, safety net hospitals that aren't doing well could learn from the ones are doing really well.

That is what we want to enable, I think, and I am going to defer -- How you use a measure like this to drive policy is an important question, but I am not concerned that there are differences there now, because it could reflect a reality that patients who

are lesser off or minorities aren't getting as good a quality care. That is possibly what is going on, and it is being reflected here, and that is what I think probably going on. The question is how can we eliminate that and do something about it constructively.

CO-CHAIR TIRSCHWELL: Risha?

DR. GIDWANI: I have a number of comments, which is probably not a surprise to anybody in this room.

I think that it is nice to see a high correlation between administrative data and medical record data, but the correlation doesn't give us a whole lot of confidence, if we see that the C statistic, the discriminative ability, is so low.

So based off of what the developers showed us, they showed a C statistic of .60. That means that 40 percent of the time their models are not able to properly discriminate or properly predict the people that actually got readmitted from the

people that didn't get readmitted.

So if the medical record model is also has that same sort of 40 percent inability to properly predict people that got readmitted versus not readmitted, then it is going to have a high correlation, but it is still indicating that this is not a well performing model.

So in saying this, I don't want to take the developers to task for this. I think that they are doing an admirable job with a very complex and sophisticated methodology, which is also a relatively nascent methodology.

A recent systematic literature review published in JAMA in 2011 looking at the ability of predictive models to look at readmission found that most of them don't perform that well. I think the highest C statistic that we found was a .77. So, of course, that is much higher than a .60, what we are dealing with here, but I think, really,

a lesson learned from the field is that this is a difficult thing to do right.

With that, these C statistics are concerning to me, and the developers have numerous times said that this lack of predictive ability, the sort of 40 percent that we are leaving on the table, is due to hospital level factors, and they are correct that we don't want to adjust for hospital level factor.

We want to illuminate a hospital level factors, and by addressing for them, we bury them. But it seems to me that the only way that you can actually reach the conclusion that it is the hospital level factors that are responsible for leaving the 40 percent predictive ability on the table is if you actually do models that include these hospital level factors, and then test the models that have the hospital level factors against the models that don't have the hospital level factors, and then you are really going to

understand what is the influence of these hospital level factors.

I think, to make conclusions in the absence of evidence is an exercise in using anecdotes to arrive at conclusions, and at this level I am very concerned about that.

It also becomes even more concerning to me that we are now going to be expanding this measure beyond Medicare to an all payer population. My concerns are strong when we are just looking at the Medicare population. They become amplified when we have a larger patient population to which this measure applies.

DR. BERNHEIM: Could I respond?

CO-CHAIR TIRSCHWELL: Sure, go ahead.

DR. BERNHEIM: Hi, sorry to not be there in person. This is Susannah Bernheim, and I just want to step back. I know I have said these things before, but I think there is a very fundamental issue that we have to come

back to, which is this is not a predictive model.

We are not aiming to give somebody a risk score that says, if your patient comes in with these factors, this is their chance of readmission. The C statistic is one and not the most important measure of the performance of these models. If we were attempting to predict every patient's readmission risk, we would not do a terrific job at it, and that is not what we are trying to do.

We, I think, have -- and I think this was probably naive -- oversimplified the idea that there are, quote, "hospital level" factors that affect these readmission rates, and I don't think that I could sort of put in teaching status and PCI status and suddenly give you a measure that explained all of the admissions risks.

What our understanding of what contributes to readmission risk is that it is actually pretty complex, which is why it is

important to use an outcome measure in this case. If it was as simple as handing patients discharge instructions, then the process measures would be sufficient.

What we are learning from an emerging literature, which is new, is that there are, in fact, very good evidence that hospitals that put systematic programs into place -- you know, Komen's Care Transition Project being a great example -- reduced admission risks and reduced patient readmission, including stroke patients.

So I don't think we can build you a model that tries to account for all of the web of things that hospitals do around patient education, around communication with outpatient providers, around appropriate next site of care, and making sure that patients have the support they do.

It is a complicated thing that hospitals are working very hard on right, but we have more and more evidence, including good

trials that are small and early, that there are things that hospitals can do that improve the outcomes for patients.

CO-CHAIR TIRSCHWELL: I am not sure who was next. Salina, then Michael.

DR. WADDY: I certainly agree that readmission is very complex and there are multiple factors, but unless we really know what percentage of readmission is actually due to hospital level factors and things that they can actually change versus the activities that patients actually do such as --

CO-CHAIR TIRSCHWELL: Noncompliance.

DR. WADDY: -- yes, noncompliance, not filling their medications, not going to their physician, not having a physician who can see them within a 15-day period or something like that, then it is really difficult to understand how they are going to use the information gathered from this measure to ensure that there is quality, because you

can get -- you may be getting quality care within the hospital and through the transitions, but it may be even more of a patient factor, and is there any way to really tease at least some of that apart so that the people who should be dinged for providing poor care -- that that actually happens.

CO-CHAIR TIRSCHWELL: Michael, Daniel, Risha, and Peter.

DR. KAPLITT: The major concern with this for many of the outside groups, which is what begat this re-review, was the risk adjustment strategy, and that is what we have been talking about a little bit.

American Heart, I think, put it fairly clearly and, I think, somewhat convincingly. The issue of the C statistic, in my view, is not so much misunderstanding the nature of what this measure is. We know it is not supposed to be a predictive measure, but it does somewhat reflect the quality of the risk adjustment strategy. At least, that

is my understanding of the way we have discussed this in the past.

The response about the concern about all these socioeconomic factors not being brought in and how important they may be in emerging literature, etcetera, was -- at least, the written response was, well, we think the hospital factors are of primary importance, but that is not necessarily proven.

The second thing is that in the answer -- and this was again stated just now on the telephone -- that if the major goal of this is to try to promote improvement and see hospitals improve their readmission rate, that is fine, but this is not a change statistic that is being looked at. It is an absolute number.

The measure is not to look at the change in readmission rate over time. It is an absolute number. So hospitals that are at a disadvantage are going to be reported as

such. It is not going to be reported -- So if a hospital is at a disadvantage and they wind up improving dramatically, but they are still below another hospital that is not at a disadvantage, they will still look bad, because that is not how this is going to be reported.

Then finally, the issue about, well, we are not that worried about this because of the fact that there is great variability among the economically disadvantaged hospitals.

I would ask (a) is that degree of variance the same as in the nondisadvantaged hospitals, because if the degree of variance is difference, it again would suggest different factors in the different groups, even if there is great variability. Great variability is not necessarily a comforting factor unless there is equal variability among all the groups. So I would ask the developer, is there equal variability?

Then I still think that all of these concerns are valid, and I don't see that they have been sort of well addressed.

DR. LABOVITZ: This is a redo of a redo. I think the reason is that some of us have played John Kerry. We were for it before we were against it, and that speaks to how incredibly difficult this measure is.

I have struggled with it. I think everybody around this table has struggled with it. I think what I have not yet understood in all of our conversations is why we are in such a rush to put an NQF stamp of approval on a group of studies that are nascent in an area that is evolving, in a place where we don't even have the data on how hospitals in caring for disadvantaged populations might be dinged for the nature of the work they do, and not for the quality of it.

I just don't see why we have to rush to approve this. This is excellent work, and I am not allergic to using hospital

readmission as a quality measure, the way I am allergic to using death as a quality measure in stroke.

I think there is real value to be gotten ere, but I think we could harm ourselves badly if we rush too quickly, and maybe next year isn't too late, or the year after that.

DR. KRUMHOLZ: This is Harlan Krumholz. Can I just say a few words?

CO-CHAIR TIRSCHWELL: Yes, go ahead.

DR. KRUMHOLZ: Thank you. So here is the urgency. The rates are above 20 percent. For years and years in this country, people have ignored the fact that one out of four or one out of five patients who leave the hospital have such a catastrophic event happen in the next 30 days that they require an acute hospitalization again, even though anyone who has just been in the hospital has a natural aversion to want to come back in the hospital.

Every time you look deeply at the transition process, you realize that it is extraordinarily flawed. We are so poor at this. I don't know any hospital who began to do a deep dive into their transition process who doesn't recognize that they don't reconcile the meds correctly. They do a poor job on education. They are not communicating well. Their discharge summaries aren't ready on time. They are not getting to the right people. They are having trouble making appointments.

Anyone who has had a recent relative in the hospital or has been unfortunate enough themselves to be in the hospital knows how baffling this process of transition is and how broken the current system is.

I understand the pain of the hospitals who are concerned that they are being disproportionately discriminated against, because they care for vulnerable

populations, but the overlap in the populations plus the fact that the absolute differences here we are talking about are minuscule compared to the overall 20 percent, 22 percent, 17 percent that we are seeing for all these different conditions tells you that, if hospitals can work together with their community, can fix these systems, can reduce their risk for patients, make safe passage possible, have people be confident in what is happening as they move from the inpatient to outpatient, then there is a large -- There is just no way there is not a large opportunity here to make this better for patients in America.

We can wait this year. We can wait next year. We can wait five years. We can wait 10 years, but God forbid that anyone you know gets in the hospital and has to manage this transition from inpatient to outpatient.

Stroke patients are particularly

vulnerable. They are weak, tired, often with disabilities, and they are in a poor position to manage this transition, as we see from their readmission rates.

By putting this up, we are going to get attention on this problem, and we are going to get it focused on this particular vulnerable population, and I challenge the hospitals to take ownership of the part that they do own, because it is enormous. The gaps are enormous.

One thing about the C statistic. We purposely are profiling hospitals. Just to put it in perspective what Susannah was saying, if I really wanted to predict, I would use all the information on hospitalization up until discharge. Five times zero is admission, because we are trying to say when the patients come to the hospital, what are they like. We are not including any information from the hospitalization.

That by itself disables the

predictive model. When I look at all the things that go wrong for patients as they make the transition from inpatient to outpatient, it is not surprise to me that the severity of disease ends up not being a very important predictive factor.

These patients are vulnerable and susceptible to a wide range of things: Infection, kidney problems, in addition to the reason that they were initially admitted. So we need hospitals to open their eyes to see a holistic view of the patient.

If we are truly patient centered, we are realizing that people are suffering every day because we are waiting for the perfect measure; and if you want to get those patients and let people go for some other areas, okay, but I am thinking that this is a group that we want the nation's hospitals rolling up their sleeves and saying how can we make safe passage for people who are admitted with stroke.

So that is why the urgency. You asked why the urgency? Because every day patients are facing a one in five chance of getting back to the hospital within 30 days and terrible things happening to them in the 30 days, which I believe, to a large extent, that risks can be modifiable, if the hospitals working with their community address them, and if -- We may be at the margin. There are things that are going to need to be addressed with regard to vulnerable populations, but every time we looked at STS, the differences are minuscule compared with the overall risk.

We have been talking to MEDPAC, too. The policies are likely to evolve. The current -- If you are above the average, below the average, yes, I don't like it either. I don't think that was a good policy. But that is not our job.

Our job is to say can we put out a measure that is the best possible measure that is going to draw attention to this issue, and

is ultimately going to help patients around this country by getting people to invest in having benchmarks to look at.

That is what I think. That is why this measure is good enough to move forward, and I hope that you guys will see it in that perspective.

CO-CHAIR TIRSCHWELL: Thank you, Dr. Krumholz. I guess the thought that came to me during your comments there is that, if we really want to spur hospitals to action, I am wondering why we don't put together an evidence based intervention at hospital discharge that every patient has to have marked off to get credit for.

That would be a direct connection to action for every patient, and I guess the answer is maybe that evidence based intervention doesn't exist. But there are a number of other comments around the room.

Risha and then Salina.

DR. GIDWANI: Thanks. I think,

sort of in response to what the developer was just saying, it seems as though the developers are coming back to care coordination and transitions of care as important hospital components, and I think it is very important to reduce readmissions. I don't think that you will find a lot of disagreement in the room here.

It is just really a question of whether this measure is the best way to do that. If the developers strongly feel that there is an evidence base regarding care transitions and care coordination and the like, then I would suggest that there is an opportunity to develop measures in those areas.

This measure in particular isn't getting at that nor is it going to give hospitals information to say care coordination is important, you should focus on that. I think, really, this is about saying this is a valid prediction, if we are looking at better

than expected, worse than expected, as expected. That expected is inherently a predicted variable.

I have looked at a lot of the model diagnostics, and I have read the methods reports thoroughly, and there are a variety of other statistics I looked at, like there is a lack of fit, amongst others.

I keep coming back to the C statistic, because it is a highly informative model diagnostic, and in this case it is also a worrisome diagnostic.

I would also like to point out to the committee that, when we look at the mortality model, the 30-day mortality day that has now been withdrawn, that C statistic was, I think, .72, .77, something in the mid to low .7.

The Fonarow, et al., article found that including the NIHSS in this model, improved the C statistic by, I think, .7, and that resulted in 26 percent of hospitals being

reclassified from better than expected to expected, expected to better than expected, worse than expected. Across those three categories, there was a movement of 26 percentage of hospitals.

So this C statistic does actually have a real impact on the categories that hospitals find themselves placed in.

DR. WADDY: I would like to support what Risha just said, that a lot of what was discussed was really the transition from hospital to the community, and there is a huge gap there. If there were ways to smooth that transition, and actually there are ongoing studies, one that should be unblinded in the next couple of months funded through NINDS, in order to help that process.

As well, there is a sort of intervention with the guidelines. Some hospitals are actually using those to really change how they perform the discharge instructions.

So, to me, having specific measures for the facility would be more appropriate than something that is more of a grab bag that is a bit after the fact and a messy measure.

DR. SCHMIDT: I just want to remind people, when I did the call on the training for this, one of the things that the NQF staff tell us is that they like outcome measures, and this is a proxy outcome measure. One could say that the hospitalization is actually processed, but it is a proxy for a bad outcome.

So there is some value in -- and whether this is exactly right is -- I didn't vote on it the first time. So I am not voting on it this time, but whether there is some value in it is up to you guys to discuss. But it is valuable to have outcome measures in the mix and to allow clinicians to come up with solutions that fit their care context to achieve improved outcomes.

MS. SUKO: I think the question -- What I am about to say echoes much of what you are going to say -- is that we don't know what those system level factors are. I think the question is: In the absence of an outcome measure, will we as an industry learn what those measures are and, if we are to approve this outcome measure, can we be comfortable with the fact that we are going to learn what those are. Then do those benefits outweigh the risk of the potential negative interpretation or the penalties of hospitals?

MS. VAN DE KAMP: I think I brought this up the last time, but just some clarification. We already have a rehospitalization measure in effect today. Correct? For hospitalization? Just overall, right?

DR. BURSTIN: We have endorsed a hospital-wide all-cause readmission measure. It is not in use. CMS could clarify that.

MS. VAN DE KAMP: Okay.

DR. GIDWANI: There is heart failure, AMI and pneumonia, though.

MS. VAN DE KAMP: That are in use. So I guess there is already measurement, and it is diagnosis driven, to some degree. Is it measurably different than this measure?

DR. DRYE: This measure is very similar to the AMI, heart failure, and pneumonia readmission measures that are currently publicly reported.

CO-CHAIR TIRSCHWELL: Have we learned things from the public reporting of those other measures that have led to dramatic benefits for patients in terms of readmissions and the processes or factors that are involved in readmission?

MS. DRYE: Harlan may still be on the line. I think we have generated a lot of focus on this transition for patients, which -- I appreciate the comment about outcomes measures. There isn't going to be one solution, that there is one size fits all.

There is not going to be one checklist. There is not going to be one --

DR. KRUMHOLZ: We have hospital -- Virtually every hospital in the country right now has now focused attention on readmission, and we have yet -- I mean, this is a big complex problem. If you do one fix, it is not enough.

If you create timely discharge summaries, it is not enough. You have to make sure that they are addressing the right information, that they are getting to the right people, that the patients are seeing the people who have in their hands a discharge summary, and that is just one little small aspect of this. But I can tell you, I mean, I have spoken to hundreds and hundreds of hospitals over the last two years, and people are with enthusiasm recognizing their deficiencies in this area and are redoubling efforts to --

First of all, before you get into

fancy new innovative solutions, the blocking and tackling has just not been done. I mean, just the communication and do people have a place to go, and they know what is going on when they leave them.

Again, any of you who have had someone -- not just caring for someone, but actually having it where you are seeing through the patient's or family's eyes what it is like to leave the hospital, you know.

These are efforts in every hospital that they are starting with. You know that we have been public reporting for a couple of years. It is going to take some time to infiltrate and change the way people are thinking, especially since we all train just to get people out the door. But I can tell you, there is a sea change in the country now compared to when we first started proposing these and since these have been publicly reported.

I can't report back the results

yet that I am happy with, but I can tell you the efforts there and around the country. There is a lot of creativity being applied to how best to do this.

The other measures are out. So, really, it is a question of whether you think you want stroke to be out of that group, but that is what is happening.

CO-CHAIR TIRSCHWELL: Daniel.

DR. LABOVITZ: This is a response, I think, mostly -- I don't know his name, but the voice in the sky.

DR. KRUMHOLZ: Harlan Krumholz, sir.

DR. LABOVITZ: I welcome your passion on this, and I think your comments on how important this is as an area to target are very meaningful. I think we all see the point. It really does matter, and I can certainly say as a provider at the beside that we do a lousy job of paying attention to this.

What I am questioning is whether

this measure should be approved by the NQF because it meets the standards, and I would suggest that we have rejected other measures which were very meaningful to us and important, and we saw the point, but the measure wasn't ready.

I think we have heard good feedback today to suggest that this measure maybe isn't ready either. There are also, I would suggest, alternative means to using an NQF measure to drive change.

CMS is experimenting now with developing a reimbursement system where the hospital is responsible for all care, not only during the index admission but for 90 days afterwards or 30 days afterwards. It will be rolled out to a subset of hospitals. They are going to play with it.

That is going to drive change within the hospitals. The hospitals will be reimbursed based upon their performance. That is going to drive change, too, and maybe we

need to see those things emerge before we start putting an NQF stamp on a measure just because it seems like it is really important.

CO-CHAIR TIRSCHWELL: Peter.

DR. SCHMIDT: I just want to point out that the ones that we rejected that we all cared about were all process measures, and there is a separate -- you know, on the form there is a separate standard, if you accept that this is an outcome, which is questionable. But if you accept it is an outcome, there is a separate standard that is very different from the way that we assessed all the process measures.

CO-CHAIR TIRSCHWELL: Developers want to respond at all to anything further?

DR. DRYE: I just -- I promise this is the last thing I will say about the C statistic, but I think it is really important to say.

With all due respect to the committee member bringing this up, the C

statistic is appropriate for this measure and scientifically valid. It is aligned with C statistics, with measures previously approved by NQF, including a PCI readmission measure developed off of registry data. It is completely aligned with the C statistic in a readmission measure we just built with the Society of Thoracic Surgeons for CABG readmission using their registry data.

You cannot -- If you follow the scientific guidelines for outcomes measures, which says you do not risk adjust for things that happen in the hospital, only things prior to the hospital, there is no way for readmission you can get a C statistic that is much above .6, and I have one very vocal member on this committee, but this is a fundamental scientific point about the validity of the measure.

It is completely consistent with prior approval and every readmission measure that I have seen, whether you use registry or

claims data.

Then I just want to make a comment on outcomes measures. I really like the way one of the members characterized the measure as messy. Outcomes measures are messy. They are really, really hard, but we are being called by patients and by providers to use them, because they are what matter to patients, and they are the end result of a lot of different complex processes that people point out, which we will never be able to capture process measure by process measure by process measure.

So I just appreciate the committee's deep review of the outcomes measures against the criteria that NQF has set forth as consensus based guidelines.

DR. HAN: Hi. This is Lein Han from CMS. I just want to share our experience with you. We are -- Someone asked whether CMS has a hospital-wide readmission measure. Yes, we do have one, and right now we are

conducting a dry run. It means that we work with hospitals, share the data with them, but we don't publicly report their data.

So during the dry run, we have conference calls with all hospitals in the nation, and the first time we have probably 2,000 hospitals call in, and the second time we have about 800 hospitals call in, and at each time it is about 90 minutes.

So during the call, we heard from hospitals. They want to know more about their data, how are they doing. I think people are -- At the hospital level, they are in the field. They don't know the big picture, how they are doing.

So they give CMS feedback: Can you give us more data where a patient goes, and also can you apply this measure to other populations like a pediatric population? After those calls, I just got a feeling. Hospitals want to know how they are doing.

The CMS developed measure, we see

a health care problem here, and we use the best data available to CMS and procure the best team in the nation to develop a measure. So our point about -- First, there are always outliers, but our goal is really to move the whole curve of distribution of the hospital performance to a lower mean.

If we have the best data, even we have access data, we will do it, but we can't wait to have the best data and the best measure to move our quality agenda forward. So I just want to share this perspective with you for you to understand where we come from to develop these measures, to implement these measures.

DR. GOODRICH: This is Kate Goodrich from CMS. I wonder if I could maybe say something as well.

CO-CHAIR TIRSCHWELL: Yes, go ahead.

DR. GOODRICH: Okay. I am the Acting Director of the Quality Measurement

Group. I work with Lein and others, I think, are there. So I want to say a couple of things.

I think, to address the point about why don't we use a number of care coordination related process measures, we agree that those measures are important, and we want to use those type of measures. But I think there is often a potentially false assumption that there is a direct correlation between how a hospital or a provider performs on a process measure, how it is directly related or not directly related to how they perform on the outcome measure.

In fact, we know from an emerging literature that there often is not that correlation, as one might expect there would be. So the use of process measures to drive improvement, we think, is nowhere near as powerful as the use of these patient centered outcomes measures.

I do feel a little bit -- I want

to echo what I think others have said. I feel a little bit like the patient has gotten a little lost in the conversation, and for us that, obviously, our primary concern.

We know that there is really nothing in terms of measurement that is going to focus the country like a laser on a serious quality problem like use of these outcome measures. But I will say we have heard from many stakeholders the concerns about the SES adjustment and race adjustment, and we hear those concerns loud and clear.

So we have started to think within our programs about how we could change some of our implementation policies. There is nothing absolute or final I can tell you about this now, but because we have been hearing this concern for a long time now, we are starting to think about how we can implement our programs or modify the way we implement our programs to address some of those concerns.

So i do want the panel members to

know that we, too, are concerned about those issues, although we agree with the NQF policy of not adjusting for those factors. So we are trying to find other policy related ways that we could address some of those issues.

CO-CHAIR TIRSCHWELL: Risha, do you have an additional comment?

DR. GIDWANI: I will just start by saying that I do applaud the developers for their effort in this regard. I think they do have a great team, and they had a sophisticated approach.

I think, you know, they are just hemmed in by the limitations of the field in general, and they are correct. The C statistic of this model is in line with the AMI, the heart failure, the pneumonia models, but that, to me, is not a reason for endorsing this measure. It is more of a reason for me to be concerned about those other measures.

I will also point out, the concerns about the C statistics -- I am not

the only one that has them. There is systematic literature of readmissions models in JAMA looking at all different kinds of hospital readmission models, generally 30-day, came to the same conclusion about predictive ability of models, and essentially said that better approaches are needed to assess hospital performance.

My concern is really that -- I guess I have another point, and then I will bring up my last concern. My other point is that we didn't pass or I think we were neck and neck and now it has been withdrawn for the mortality model, and that had a much higher C statistic, and people were concerned about the fact that the NIHSS wasn't included in that model, because when it was included, the C statistic and model performance improved.

So as a matter of just internal congruence on our own panel, I just wonder about not endorsing a measure that had a higher C statistic and endorsing a measure

that has a lower C statistic. It does worry me.

Then really, my major point of concern is what is the harm that could happen. I want to make sure that the patients are being given the appropriate care. My worry is that, if we have a measure that doesn't properly reward hospitals for good performance and disincent hospitals to avoid bad performance, that we could end up with a situation where certain kinds of patients are just refused for admission into the hospital because they may be at high risk of subsequent readmission.

CO-CHAIR TIRSCHWELL: Mary, go ahead.

MS. VAN DE KAMP: I think Kate might have alluded to this with her discussion about implementation. I think the fear is not that we measure or that we look at it. It is that we too quickly go to payment impact from that.

I think, if there was a way for us to look at these, because I do think -- I am in the field. I see what those outcome measures in the other diagnosis have done to the analysis of transitions of care, and it has been very impactful. But what it has also done, I am fearful, is that not everyone looks at it that way, and there are others who prevent readmissions or are concerned about readmission because they don't want the rate to impact their financial payment.

While it is two percent the first year, which is significant and not insignificant, I think one recommendation -- and, I think, a less fear factor for those of us in the room -- would be if the payment didn't follow so quickly on the outcome measure, so that we could really look at the quality outcome and then, after a certain period of time, determine did it measure the right thing? Did things improve and, therefore, it is a requirement for payment.

I think the fact that we have been struggling more than anything is that there is a very quick correlation to either a public reporting of poor outcomes -- so publicly the consensus drops or, secondly, there is a reimbursement change. But I absolutely see what you are saying in terms of people really, really looking at -- I mean, the skilled nursing arena.

I know that hospitals are looking at post-acute discharge, because they want to make sure that they are going to send to a quality nursing center that doesn't send their patients back.

So I get all that, and I think it is very important. I think the payment and the public analysis is there.

CO-CHAIR TIRSCHWELL: Sure. One second, Daniel. Go ahead.

DR. BURSTIN: This is just one process point. So the endorsement process is really about the measure properties of the

measure. It is not in the purview for us to really be, at these tables, talking about applicability of measures for different purposes.

That is really what our measures application partnership is about, and they would likely to have an opportunity to consider which accountability applications might be most appropriate for additional readmission measures.

So just a reminder. We have got to stick this to the actual measure itself, not the issues around payment and public reporting.

CO-CHAIR TIRSCHWELL: Daniel.

DR. LABOVITZ: This is a comment on or maybe a question on whether this is an outcome measure. Peter Schmidt's comments point out that this is a fundamentally different sort of beast than a process measure are very well taken, and I think the messiness of it much appreciated and, I think, important

to consider as we move forward. But I am not sure that readmission is an outcome.

I am not sure it should be judged the same way you might judge length of life or quality of life, which I think are meaningful outcomes, or even reoccurrence of a dread disease. Readmission has many, many factors contributing to it.

I know that there is some data to suggest for a heart failure readmission that people who come in more often have a higher quality of life. The higher readmission rate means that they are actually being cared for better.

So I don't feel like I understand this. It is, dare I say it, squishy.

CO-CHAIR TIRSCHWELL: Last word, huh. Any other comments? I guess a question that needs to go out to you all is have you seen and heard things over the phone calls and today to suggest that you want to revote on this measure?

I guess, if we don't revote, then the last vote stands, which means it would not be endorsed. I don't know what the process point is here. I am thinking it is like the exception. If any one person thinks we should revote, we probably should go ahead. Does anybody want to suggest that we revote on this measure?

I am not seeing anybody suggesting that. So then the last vote stands, which I believe was 10 votes for and 12 votes against.

The next thing on our agenda is member and public comment. Arnica, can you open the lines to public comment.

OPERATOR: At this time, there are no questions.

CO-CHAIR TIRSCHWELL: Thank you very much. Then Karen, do you want to talk about next steps and committee timeline?

MS. JOHNSON: This is where I get to hand it over to Suzanne.

CO-CHAIR TIRSCHWELL: Okay.

Suzanne, thank you.

MS. THEBERGE: Thanks, everybody, for your time today. Next steps are going to be the same as after the last meeting. We will write up a draft report and put that out for comment.

Right now we are estimating that is going to go out around October 31st. That is a 30-day comment period. So it will close right after Thanksgiving, and then we will have our conference call in early December for you to discuss those comments. I think that call is scheduled for December 10th, but I can't remember at the moment. So I will follow up with you all by email about that next week.

Then after that comment call, we will do the same thing we did last time. We will discuss the comments received, if you need to revote on anything, and then we will go to NQF member vote, CSAC and Board approval.

So that is next steps for this project.

CO-CHAIR TIRSCHWELL: And just so I am clear, the things that will go out for public comment are those that were approved?

DR. BURSTIN: We want to get public comment on the things that are approved as well as disapproved in case people want to bring other evidence or information to bear. We will bring it all back to you for transparency.

DR. GIDWANI: I just have a question about the time endorsed -- time limited endorsed measures. I am not sure how many of those we had, but if we did have any, is the committee going to be seeing any data about their reliability or validity or what is the process for those?

DR. BURSTIN: Typically, the process is that the measure testing results go to our Consensus Standards Approval Committee. So they look at all the testing results.

DR. GIDWANI: Then also in terms of what happens from here, we have endorsed certain measures. Does then the higher level CSAC need to also approval those before they become formally NQF endorsed?

DR. BURSTIN: Actually, the CSAC and then the NQF Board ratifies that decision. The SAC is talking about the first phase next week, Monday actually. Then it will go to the Board. It will be endorsed shortly.

DR. GIDWANI: And what are their requirements for endorsement? Is there a likelihood that what we endorse does not come to bear?

DR. BURSTIN: Generally, most of the decisions remain endorsed, and I think the difference is that the CSAC and the Board often has a very sort of high level view of the entire portfolio, and at times we do get different perspectives on importance of measures.

This group is so heavily clinician

oriented. When you get into a group that is more balanced of consumers, purchasers, and all stakeholders, sometimes, particularly around importance, some of those issues may bubble up differently.

CO-CHAIR TIRSCHWELL: I would just like to personally thank everybody for all the time they have put in on this, and hope to see you again sometime soon.

CO-CHAIR KNOWLTON: And before you all leave, I especially wanted to acknowledge the NQF team. There is so much work behind what we do that these guys do. I want to mention that.

DR. BURSTIN: We have to acknowledge our Chairs who did yeoman's work, I think, over two meetings. So thank you.

(Whereupon, the above-entitled matter went off the record at 2:01 p.m.)

H

T