NATIONAL QUALITY FORUM

\+ + + + +

PALLIATIVE CARE AND END OF LIFE CARE STEERING COMMITTEE

\+ + + + +

THURSDAY
JULY 21, 2011

\+ + + + +

The Steering Committee met at the Capitol Hilton, 1001 16th Street, N.W., Washington, D.C., at 9:00 a.m., R. Sean Morrison and June Lunney, co-Chairs, presiding.

PRESENT:
R. SEAN MORRISON, MD, Co-Chair
JUNE LUNNEY, PhD, RN, Co-Chair
RUSSELL ACEVEDO, MD, MD, FACP, FCCM, FCCP, Crouse Hospital
EDUARDO BRUERA, MD, FAAHPM, The University of Texas MD Anderson Cancer Center
DAVID CASARETT, MD, MA, University of Pennsylvania School of Medicine
ROBERT FINE, MD, Baylor Health Care System
RICHARD GOLDSTEIN, MD, FAAP, Dana-Farber Cancer Institute
SARAH HILL, MA, Ascension Health
PAMELA KALEN, National Business Group on Health

NAOMI KARP, JD, AARP Public Policy Institute
MICHAEL LEPORE, PhD, Planetree
SOLOMON LIAO, MD, University of California, Irvine
STEPHEN LUTZ, MD, Blanchard Valley Regional Cancer Center
HELENE MARTEL, MA, Kaiser Permanente

NAOMI NAIERMAN, MPA, American Hospice Foundation

DOUGLAS NEE, PharmD, MS, OptiMed, Inc.
KATHLEEN O'MALLEY, California HealthCare Foundation
TINA PICCHI, MA, BCC, Supportive Care Coalition
TRACY SCHROEPFER, PhD, University of Wisconsin-Madison School of Social Work
DOUGLAS WHITE, MD, MAS, University of Pittsburgh, Department of Critical Care Medicine

NQF STAFF:
  HEIDI BOSSLEY, MSN, MBA
  HELEN BURSTIN, MD, MPH
  ERIC COLCHAMIRO
  CAREN A. GINSBERG, PhD
  LINDSEY TIGHE, MS

ALSO PRESENT:
LAURA HANSON, University of North Carolina Chapel Hill*
DALE LUPU, American Academy of Hospice and Palliative Medicine
CAROL SPENCE, National Hospice and Palliative Care Organization

MARTHA TECCA, Deyta
JOAN TENO, Brown Medical School
NEIL WENGER, RAND*

*Participating via teleconference

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

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

9:01 a.m.

DR. MORRISON: Good morning, everybody, and welcome to our second day. Again, thanks to everybody for a really, really productive session yesterday, after what several people told me might have been a little bit of a difficult start.

I do want to reassure you that I think today -- he said, knock wood -- is going to go very smoothly. We have -- Oh, don't laugh. We have our measure developers available to us to answer any questions. Most of them present at the beginning measures, and the preliminary scoring looks pretty straightforward. So I think we are in very, very good shape.

Just very, very quickly on today's agenda, we are going to spend the morning on reviewing one, two, three, four, five measures. We are going to break for lunch. We are going to have one, two, three, four, five -- four measures -- three measures in the afternoon. Then there will be a brief time to discuss gaps, after which you guys have seen everything here, and then for us to identify for NQF what specific gaps are there, what we didn't see, and what are the future opportunities.

I need to apologize to the group. I am going to be leaving about 12:45-1:00 o'clock to get a flight home. I have to get my two kids to camp in Maine by tonight, and this is the only way to do it. Unfortunately, their camp schedule was scheduled before this meeting. So June is going to take over and run everything for the last part of the afternoon.

Just a recap of yesterday. We reviewed eight comprehensive -- or eight measures. We approved seven of them, and we tabled a number of measures for further discussion.

Just for the group, June and I and

Lindsey, Karen and Heidi will be talking with Dr. Earle next week by conference call. We, hopefully, will get answers to all the questions that you guys put forward to him. We will circulate that back to the group.

Then we will reconvene by electronic voting to go through the ASCO measures and readdress those and, hopefully, the answers that Dr. Earle gives us will give you guys the confidence and the ability to actually vote on those in a meaningful manner.

Make sense? Sir?

DR. BRUERA: Sean, I wanted some clarification. Dr. Earle has changed his role since this was done. So one of the questions would be: Who is the ASCO person who could further strengthen some of this work, because if he has a different role and a different scope, some of the very important questions for this that I personally, coming from the cancer perspective, feel are the most important outcomes, might not be strengthened.

So what would be the process to get the sponsor institution like ASCO to be able to successfully address the comments that we have to strengthen these measures for the next application?

DR. MORRISON: You are exactly right, and that is one of the issues that Lindsey and I were talking about before the meeting. Let me say it this way. There was an opportunity to identify somebody from ASCO before the meeting, and Dr. Earle didn't feel that there was anybody at ASCO who could step in for him.

I think that what I would like to do with the NQF staff is reapproach that very closely, because as you said, ASCO will be the steward of these. ASCO is the steward of these measures moving forward, and I think that, given the fact that they have taken on that responsibility, my perspective is that it would be very helpful to have somebody from ASCO step up to the plate, since they will be stewarding these, and Dr. Earle is back in the great white North where he is really not connected to the American Society of Clinical Oncology.

So I agree with you completely, and I think it is something that we, with the NQF staff, will try to explore really well, because I completely agree with you, Eduardo. It is critically important. There are some important measures in an important population, and we need to see if we can get them right.

MS. BOSSLEY: Just to add to it, I actually have the text on my other phone that I didn't see. There was an ASCO representative on the phone listening. So I think we have an opportunity to get a phone call with Craig and also an ASCO representative and talk this through, because I think they have data that will help with this. I think they are aware of it. I think we just need to have a few more conversations.

The other thing is I think,

because you didn't have a very robust discussion on all of the measures, we will try to have Craig on a phone call with all of you to go through those measures, and then have you vote after that, and ASCO will also be on.

So that is something that Lindsey and Eric and Karen and I will be working on over the next week or two, to figure out when we can do that.

DR. MORRISON: Everybody ready to move forward? Did I miss anything else, Heidi? Okay. Anybody have a conflict of interest or disclosure that happened between yesterday and today that needs -- Don't laugh. The last NQF session, we had to do this all over again. So just checking. It wasn't you. No, no, it wasn't you.

All right. So we are going to move to the first measure, which is 1626, patients admitted to ICU who have care preferences documented.

Do I have somebody from RAND on

the phone who can present? Anthony, do we have anybody from RAND on the phone?

OPERATOR: Not at the moment, sir.

DR. MORRISON: Was that a yes?

OPERATOR: No, sir, not at the moment.

DR. MORRISON: Not at the moment, okay. So who do I have as the presenter? Tracy. Thank you. So I have Tracy as the presenter. So I am going to open it up to Tracy to present. We could do that. Is Laura on? Is Laura Hanson on the phone?

OPERATOR: She is joining shortly.

DR. MORRISON: She will be on shortly? Okay. How about somebody from Deyta? Yes. We are going to go down to 1647. David Casarett, could I ask a favor? Could you just move a little bit this way? Thank you. Thank you, thank you, thank you.

So this is documentation of spiritual/religious concerns, 1647.

MS. TECCA: That is right. I am

Martha Tecca from Deyta.

This is a measure -- Again, just backing up briefly, what Deyta does -- I know we are the only folks that are here sort of as a vendor as opposed to as a researcher or more commonly as a steward of measures.

This is not something that we have done in this way before, but we have been working to implement as many of the standard -- what are evolving as standard measures as possible among our clients who are using them for performance improvement primarily, and in some cases in some states, they are using them for public reporting.

So we have actually captured data from lots of hospices across the country for many of the measures or different permutations of many of the measures that are here.

When talking with folks in advance, as I know many of you did, to figure out whether or not we had a potential role here, we were trying to determine which of the categories of measures that we felt were really important and in the conversations appeared important, that there wasn't data, that people were not going to be presenting the measures from a research perspective.

So we selected this one measure, because we felt -- to promote, we thought we had a reasonable amount of data from actual hospices who were using the measure, and wanted to make sure that we were able to have a real conversation about this particular dimension of hospice care that we think is important; and to the extent the NQF measures of palliative and end of life care can be looking at things that are core to those services, this felt like a missing one in the list that I had seen going forward. So that is why this is here at all.

The data that we are presenting for -- and the measure itself is one that was promoted through the PEACE Project and then the AIM project. So it isn't a measure that we designed, but a measure that had been promoted as a standard and that we have captured data for, and we are presenting you with the data from our clients over the course of the time frame.

So it is a little bit of a different -- Our application probably looks a little bit different. I don't know if you need more background than that, but we are comfortable that it shows that there is, for something that is required, documentation of a spiritual conversation, required again by the conditions of participation and, certainly, a core aim and goal of this kind of care, there is a surprising amount of variation. So we thought it was useful to have the conversation.

DR. MORRISON: Fantastic. Thank you very much. Tina, I have you as the lead.

MS. PICCHI: Great. Thank you for that background. So this is 1647, and the measure is the percentage of hospice patients with documentation in the clinical record of a discussion of spiritual/religious concerns or documentation that the patient or caregiver did not want to discuss.

As Martha indicated, there really is. Spiritual care has been shown to be a critical element of quality of life at the end of life and is of significance to 1.5 million people who received services from approximately 5,000 hospices in the U.S.

I am going to start at the end and work back. Basically, I want to tell you, of the six people who reviewed this measure, only one recommended the measure for NQF endorsement, and I think there is some particular reasons why.

In terms of the demonstration, there certainly has been pretty much of an agreement of high impact, but the demonstration of performance gap, there was a variation in the reviewers' evaluation of that.

The medium score shown in the data was 78.2 percent in 2009, 73.6 percent in 2010. I think the concern is that the body of evidence was rated very low by the majority of the evaluators. There are no known studies, and there didn't appear to be enough evidence to demonstrate that this measure is meaningful.

The data does come solely from data and the system, though it comes from data collected from 13,435 records in 2009 and 2010, and it does demonstrate a less than optimal performance across the providers.

Now the citations for evidence of impact include the NCP guidelines, which have the eight domains of care, the NQF preferred practices, and also it mentions the Consensus Conference on Improving the Quality of Spiritual Care as a dimension of palliative care that was published in the Journal of Palliative Medicine and later in its entirety in the text written by Dr. Christina Puchalski and Debbie Farrell entitled "Making Health Care Whole."

Now my comments come to you also as a board certified chaplain and someone who has directed spiritual care in chaplaincy and hospitals for the last 30 years.

It appears from my reading and the other evaluators' that one of the important elements that is missing that refers to the preferred practices for the National Quality Forum's preferred practices is the use of a standardized instrument to measure spiritual distress, spiritual pain, and to integrate those findings into the care plan.

It appears that this measure falls short of meeting the NCP guidelines and the NQF preferred practices, because a standardized instrument is not being used. Also, there isn't any indication as to who this assessment would be made -- who would be making this assessment. Is it a spiritual care professional? Is it someone else on the hospice team, and are they qualified to make this screening or is it an assessment? That is another question.

The measure is a little vague in its language, only that it requires documentation that there was a discussion of spiritual concerns or documentation that the patient or family didn't want to discuss it.

It also doesn't meet some of the other NCP guidelines in terms of ensuring that, if there isn't indications of spiritual distress or spiritual pain, there would be then a referral to a qualified spiritual care provider who would make an intervention, and then integrate that into the care plan.

So based on that assessment, I think that is why the measure was not strongly recommended by the reviewers. However, I must say as a spiritual care professional, it was very heartening to see a measure being proposed for this. So I really want to applaud Deyta for putting that forward.

DR. MORRISON: Fantastic. Thank you, Tina. That was very comprehensive. Open for discussion. Dr. Casarett, Dr. Lunney, Dr. Fine?

DR. CASARETT: I would like to lead off with a question, but it is really more of a comment or maybe a question to consider in follow-up comments.

I wonder if we are letting the perfect be the enemy of the good for this measure. I understand all of those critiques that you just shared, and I read through the proposal just now and found some of my own, but I also think back to some of the pain measures we have looked at, and many of the questions that we are asking now about this spiritual assessment, we didn't ask about the pain measures.

We didn't ask them to say who would be doing the screening. We didn't ask for a follow-up. We just said pain screening is important. You can't really manage pain unless you know it exists. Therefore, we thought it was a good measure.

So while I recognize all of the flaws that you just mentioned, I would ask us in our comments to consider whether those really are fatal flaws or whether this might be a good first step.

DR. FINE: Yes. As the one guy that voted for this, I saw this as a good first step. I like to tell the people I train that sooner or later death is not a medical problem to be solved; it is a spiritual problem to be faced.

In all, I see this as a screening tool, not an assessment tool. It was maybe not worded as well as it could be, but I echo what David said, and that was my feeling when I first read it. The evidence is not great, if you look at the literature, but they collected data on nearly 25,000 patients, actually, over a couple of years. They showed variability.

I view some of what we do as a political tool, trying to move the public debate, and I think it is important that we ask hospices -- and I would again love to see this eventually expanded into the acute care setting and palliative care teams.

We on our palliative care teams at Baylor Health Care in Texas -- you know, we do, in fact, ask everybody if they have spiritual concerns and, if they do, if they wish to discuss them, what their faith is, etcetera.

We all ask it in slightly different ways. Some of us do the FICA questions, some of us don't. Then based on that, we may call in the chaplains. As a Baptist health care system, we are swarming in chaplains, and we have got a CP program. So we never lack for spiritual professionals who can come in and help.

I think you got to ask the screening question and say it is a quality measure. So I still say yea.

DR. MORRISON: Naomi.

MS. NAIERMAN: In our research with consumers, we found two things. One is that this is something they really want to know about hospice when they are selecting a hospice in advance: Will my family member, will I, have this kind of support, if and when I need it during my hospice experience?

We also found quite a bit of variation among hospices. My sense is that it is a kind of thing that can vary in terms of resources allocation among hospices. It is kind of fudgeable versus a nurse's aide or a nurse. So I would vote for this myself, despite the technical thoughts.

DR. MORRISON: Everybody will get an opportunity to vote. Other comments, questions from the group? Tina, I see you moving there.

MS. PICCHI: The concern that I have is it would be my hope that it would be part of the initial screening assessment when the patient was brought into hospice. My concern could be that it could be hours before death that there was documentation that a clergy person was called in. So that was just another part in terms of the timing that I think could be improved.

DR. MORRISON: Naomi? Naomi, Stephen, Bob, and Naomi again. Are you up or down? Okay.

MS. KARP: This is just a wording issue, but in the numerator statement, it just says documentation of spiritual/religious concerns; whereas, it seems to clearly be talking about a discussion. So I don't know if we have the ability to just amend that, to add the word discussion of, because there could be potentially -- you know, someone might just make a remark to a staffperson, and they note it, but there is not even a discussion. So I would just want that to be clear.

DR. MORRISON: Stephen.

DR. LUTZ: Yes. I am trying to figure out how to word this. So if I don't word it well, just ignore me and move on.

I am trying to share concerns that some patients have voiced to me, if they are in a system. Say they are in a system that they know is a certain religious order, and they are asked if they have any religious concerns. I basically have patients say to me, essentially, do you think it is going to matter, I am in a Baptist hospital, and I said I don't believe in God.

It sounds stupid, but I have actually had this said. I am not trying to be the unintended consequences guy for every question, but I have had this concern brought to me before. I don't know how else to voice it other than to say some people take even the question a little bit differently than maybe we would intend it.

DR. MORRISON: I have got Bob and

Stephen. Every committee needs an unintended consequences guy. So thank you for filling that role.

DR. FINE: Tina, I just wanted to respond to something you said. I didn't read it and interpret it that it meant it had to be a CP provider, a certified chaplain, which is just interesting. I don't know why a nurse or physician -- Again, I don't see it as an assessment. If somebody sees this as an assessment tool, it is not. It's lousy.

There is no assessment there at all. It just says was there some kind of discussion. Was the issue raised? That, to me, doesn't have to be done by a chaplain, and I wouldn't worry about the unintended consequence of, oh, my god, they have just been enrolled in hospice, they are dying, and the chaplain hadn't shown up.

DR. MORRISON: Tracy?

DR. SCHROEPFER: I think the other concern would be that spirituality wasn't defined. I think that people struggle and think that spirituality has to do with religion, but it doesn't necessarily. It can, and it cannot. So that oftentimes social workers are doing the spiritual, because people are talking about meaning and life, which is spirituality. How do I -- What legacy will I leave behind? Where do I fit into this world? So those kind of questions.

So I think that I was more concerned that it didn't define it well enough, such that if it is not the chaplain or if it is not religious, that then it would not be counted, and that would concern me.

DR. MORRISON: I assume, Bob and Stephen, you are okay. Right? It's okay. I just want to make sure before I move forward. I know it is early in the morning.

Other comments, questions, before we move? I figured you had said everything, Tracy. Other comments or questions before we move to voting? Are people comfortable?

Okay. We are skipping the first question again. We are skipping the first question again, because we agree on the importance, and we get an extra 45 seconds doing that all the way through the day.

MR. COLCHAMIRO: On the performance gap demonstrated by the measure, 12 high, four moderate, four low, and zero, insufficient evidence.

DR. MORRISON: I am sorry. I should be reading these. This is evidence or outcome, importance to measure and report. Everybody, try again. I am going to make Lindsey sit in the middle of the circle here. We could all go around one by one. All right, everybody try one more time.

MR. COLCHAMIRO: Six, yes; 13, no.

DR. MORRISON: Okay. Are there appropriate -- Is there a quantity of studies in the body of evidence to support it?

MR. COLCHAMIRO: Two high, 11 moderate, 7 low, and zero insufficient

evidence.

DR. MORRISON: Importance of the measure to report? Sorry, quality, quality, quality. It is way too early in the morning.

MR. COLCHAMIRO: Three high, eight moderate, eight low, one insufficient evidence.

DR. MORRISON: Okay. Consistency of results across the body of evidence?

MR. COLCHAMIRO: Four high, nine moderate, six low, one insufficient evidence.

DR. MORRISON: Scientific acceptability of the measure properties: Reliability? Of course, we can go back. A very good point, thank you.

MS. BOSSLEY: I just have a question as well, because the evidence -- The provider didn't hear it again. We will be writing all of this up. So it will be helpful to know. The evidence that is provided in here is based on -- I think it is the NQF preferred practice that we have.

There is very little else provided, but all the ratings that I saw here are showing that the quantity is moderate or so, the quality is moderate or so. So I just want to make sure that either you are voting based on something that we don't have in front of us, and your own expert opinion, which is fine. I need to know that, so we can document it, or we just need to probably have a little more discussion on that. It would be helpful.

DR. MORRISON: David?

DR. CASARETT: I was curious about that, too. I can give you my response, which was, in thinking about the evidence, I was thinking not just about what was in here, but about the literature that exists that suggests that patients do welcome the opportunity to have these sorts of discussions.

DR. MORRISON: Other comments or thoughts, because I think Heidi's point is very well taken. There is very little evidence supported, and I would add that, remember, that the NQF framework is evidence based, but certainly a substantial component of the framework is expert opinion, and I think this is one of the key issues where there was not strong empirical data to support it, and there was a lot of -- this was based upon the expert opinion issue.

So as Helen keeps reminding me, expert opinion does count as evidence, but just make sure that what you -- and Heidi is nodding. Just make sure that that is the point. So I have got Naomi, and then Kate.

MS. NAIERMAN: Just quickly, to repeat but with a little bit more information, that in our study with consumers this was a very valuable piece of information that they wanted to know, in a public report, by the way. That was the reason for our research.

If you would like a link to our published report, I would be happy to direct you there.

DR. MORRISON: Kate?

MS. O'MALLEY: I guess my question would be, for the purpose of this particular process, is it helpful to hear people say, in my expert opinion, I think this is very important; because I would say, in 20 years of clinical practice in community settings, I have found this work to be very significant to patients and families.

DR. BURSTIN: And, actually, just to remind you, the way the Evidence Task Force has set this all up for us, obviously, we went through the whole issue of quality, quantity, consistency.

There is specifically an exception to the empirical body of evidence, actually more so for -- I guess for any kinds of measures, -- where expert opinion systematically assessed that benefits to patients greatly outweigh potential harm. So if that is the direction you want to go, you just want to document that you are clearly going in that direction.

DR. MORRISON: June?

DR. LUNNEY: I was just going to comment that, in case you didn't read the paper this morning, Harvey Chochinov made the USA Today with his study. So that says somebody is interested.

DR. MORRISON: Doug?

DR. WHITE: Just then for clarity of how we should be voting on these, would we be voting according to the evidence in the room, and then there is a separate space where we say, notwithstanding this evidence, that we think, based on expert judgment, blah-blah-blah?

DR. BURSTIN: Yes. I think you all voted this sort of in the moderate range, some more lows perhaps than we are necessarily comfortable with, because why I think we will qualify in the report that you continue to evaluate the measure, because there was a consensus on expert opinion that the benefits outweigh the harms.

DR. WHITE: But is that where you want us to go?

DR. BURSTIN: I think you should -- If that feels right to you, in terms of the process, yes.

DR. WHITE: This is just a procedural question.

DR. BURSTIN: Yes. I think that would be fine.

DR. WHITE: Should we vote according to that or should we vote according to what is in front of us, and then say, but we still think it should go on? Is there a later question where we can sort this out?

DR. BURSTIN: There is not a later question, but it probably would be a good straw vote, just a hand vote, just to see if people are comfortable with that approach, because you have already voted on the quality, consistency and evidence.

DR. MORRISON: Can I see a show of hands, who are comfortable with that approach, to quote the good Dr. Burstin?

DR. BURSTIN: Would you state the approach?

DR. MORRISON: All right. So I am going to have to rephrase. So I think what I am asking is that, based upon the -- We voted to move this forward with a -- There was a higher percentage of low/moderate than we have typically seen that, I think, NQF staff feel comfortable pushing forward.

What I would like to see is a show of hands who feel that the benefits of this measure far outweigh the potential of harms to patients, that it is a measure that has very, very few unintended consequences or harms for patients, and the potential benefit is great. So just a show of hands who believe that and feel that we should move forward.

So I have got a strong majority. So we can move forward? You guys are good? Okay. We will move to the next question then. Reliability?

MR. COLCHAMIRO: Three high, seven moderate, nine low, one insufficient evidence.

DR. MORRISON: Validity?

MR. COLCHAMIRO: Three high, nine moderate, eight low, zero insufficient evidence.

DR. MORRISON: And disparities?

MR. COLCHAMIRO: One high, eight moderate, seven low, four insufficient evidence.

DR. MORRISON: I was going to stop. We are going to have to have a conversation now.

DR. BURSTIN: I think so.

DR. MORRISON: So do you want me to read this one, Heidi?

MS. BOSSLEY: I think it would actually help to have Lindsey or Eric just read us back the ratings on reliability and validity.

DR. MORRISON: That would be helpful.

MS. BOSSLEY: I think it is not clear it made it.

MR. COLCHAMIRO: So for reliability, we had three high, seven moderate, nine low, one insufficient evidence.

For validity, we had three high, nine moderate, eight low, and zero insufficient evidence.

MS. BOSSLEY: It may be helpful maybe to talk a little bit more about why people voted what they did for reliability and then also validity, because I don't know that we really got into the weeds of it on that. So why don't we talk through that a bit? Does that sound good to you?

DR. MORRISON: Yes, thank you. Solomon?

DR. LIAO: So I think it may boil down to whether we are thinking this is a screening question or whether this is an assessment.

DR. MORRISON: Sure. Yes, I am

not sure that -- Yes. Let's ask, because I am not sure.

MS. TECCA: This is, again, one of those odd places where I am calling the developer. I am forwarding this measure that was developed by PEACE, that was promoted by AIM, and we have been using with hundreds of hospices. I am not the developer of the measure, but as others have described, it is not an assessment. It is simply specifically asking whether there was a general question -- whether the conversation and discussion had been had about spiritual and religious concerns, and the documentation of that, and documentation of whether or not the family refused.

It is actually very specific. It is broad, but it is a very specific definition. It is not about an assessment. It is not about a specific assessment, not about even necessarily screening. It is about the documentation of the discussion.

DR. MORRISON: Anthony, could I ask, is Laura Hanson on the phone?

DR. HANSON: I am.

DR. MORRISON: Okay. Laura, can I put you on the spot?

DR. HANSON: Uh-oh.

DR. MORRISON: Because I have heard several times that this was developed by your group. I guess what I think it might be helpful to the group is -- you know, I have two questions as Chair's prerogative.

The first is a little bit about, given that you guys developed it, why you didn't submit it as a potential measure, and perhaps you could answer some of the questions that have arisen about the reliability and the validity and the scientific merit of this, given the committee's discussions about scores that were in the very low to moderate range, and help us with that? Doug, I have you next.

DR. HANSON: Okay. In the development, really, our logic in the development is precisely what has been described, that the evidence base is very strong for the importance of this domain and addressing this domain in some way in the clinical practice of hospice and palliative care.

When we put it before the technical expert panel for the PEACE Project, it was rated highly, not as high as some others, but certainly rated very high, over four on a five-point rating scale overall.

The concerns all along were really in the area of effective intervention to address concerns, which I think the discussion here has characterize well. So we ended up not putting it forward mostly related to those concerns, but we have some additional data about inter-rated reliability that says it can be reliably abstracted from medical records and, certainly, data for the performance gap, as I think has been comparably presented.

I think the discussion -- I have

been listening to the whole discussion, and I think the discussion fit completely with what we have learned in the process over the last several years of looking at this particular measure.

As I said, the rationale for us not putting it forward had to do with still an evidence gap between detecting spiritual distress and then the nature of intervention to relieve that distress. But we didn't find a lot of evidence for harms about asking, and I would characterize this as a screening question or as a screen.

DR. MORRISON: So I just want to -- So what I am hearing, just for the group, is the question was that, as a process measure, there wasn't -- there is not a clear link to an outcome or an intervention that would affect the outcome. Is that what I am hearing specifically?

DR. HANSON: Right, exactly.

DR. MORRISON: I don't want to put

words in your mouth, but I think it is a really important concept for the committee to consider, particularly given the focus on measurable interventions that can affect a process. Okay.

So I've got a lot of people with tent cards up. Let me start with Doug who was up first. Then I've got Naomi. I've got Kate. I've got June. I think that is it.

DR. WHITE: This is just a comment sort of again about the process here. So the question that had been asked is tell us about why you voted the way you did, and I voted the way I did, because I think it is important for sort of an integrity based view of NQF that we vote according to what is in front of us, and then we have an opportunity to say, notwithstanding this, here are the reasons that we are going against the data that is in front of us.

I would hate for the public to look at our ratings and see a bunch of ones and twos, and then look at what we based that on, and say this is not a credible process.

DR. BURSTIN: Just one point, though. That only applies to evidence. It has nothing to do with scientific acceptability. Scientific acceptability is truly what is in front of you, and the only question is whether you had sufficient information that perhaps -- I don't know that Laura could have provided additional information than what is in front of you, but again you have to act on what is in front of you.

DR. WHITE: Yes. I am talking more about the things around the measure itself, the reliability, the validity, etcetera, which is a central component of the evaluation, I think.

MS. BOSSLEY: And part of what we will do is -- Assuming this passes scientific acceptability and you move forward through the rest of it, what we try to do is then reflect within the notes the Steering Committee's rationale of why you did what you did, and you all will see that. But we try to frame it so that it doesn't sound like you -- You may have known that you did something that may not look like it followed the process. Happens all the time. We are very good at writing notes to reflect that you actually knew what you were doing when you did it.

DR. MORRISON: The joys of being in a public forum, Doug. I'm sorry, I've got Naomi, Kate, and June.

MS. NAIERMAN: I want to make two comments. First of all, Martha is right. She is not the developer of this measure, and we have just heard from the developer of this measure. As far as I recall, she said it was reliable and valid. It passes the test of validity and reliability.

The other thing is that I am not sure whether we are raising a higher bar for this particular measure than we did for previous measures where we just simply asked for screening and assessment, and not outcome.

There is, clearly, an intervention that could be applied if a screening showed that someone wants to discuss spiritual care. You just bring the chaplain in or whoever. So I want to make sure we don't raise the bar any further than we did for previous measures.

DR. MORRISON: Kate?

DR. HANSON: This is Laura.

DR. MORRISON: Sorry. Laura, I've got a couple of committee members first, I'm afraid, unless you are going to clarify.

DR. HANSON: I just want to correct that statement, that we have data on inter-rater reliability that shows good reliability. We do not have data on validity.

DR. MORRISON: That is very important. Thank you. Kate?

MS. O'MALLEY: I am adding my voice to the access issue. It sounded from Laura's descriptions they held back on the intervention side, but when I was listening to that, I also felt that we raised the bar a little bit higher, thinking about pain screenings. You can do a lot of pain screening, but you don't necessarily know always the quality of the intervention. So I think that is an important consideration for this work.

Also, in looking further at the documentation, it does clearly say that it documents a discussion about whether or not a question took place about spiritual concerns.

DR. MORRISON: June?

DR. LUNNEY: I just wanted to ask the NQF question, that I voted on my reliability judgment without hearing Laura.

MS. BOSSLEY: One thing that we could do is see if Laura would be willing to provide that information, and then bring it back to the committee. Then you can revote, and we could probably even do that essentially on the next phone call that we have with all of you. So that would be the other option.

DR. BURSTIN: It still is a concern, though, because measures at low validity don't go through. So if there is truly no validity testing, that is really where it comes down to.

DR. MORRISON: Yes, and I do think that is a very -- I am hearing that as an important point. David, did you have a comment? I'm sorry.

DR. CASARETT: No. I was just going to make that same request, but although I certainly believe Laura when she said it was reliable, I think we need to see the numbers.

DR. MORRISON: Yes, and I do think -- I'm sorry, Solomon? No, that's okay.

Heidi, I am not quite sure how to proceed here, because I think as a committee we need some information from Laura about the scientific reliability, but what I am also hearing is that we have no validity data at all. I think, from -- Let me frame this.

From a scientific perspective, it is very hard to push through a measure that has no validity data behind it.

Let me rephrase that, because Helen is saying it. It is impossible to put through a measure that has no validity data behind it, and I am not quite sure where we go from here.

DR. BURSTIN: I would actually suggest that it makes sense to actually allow the developer to speak with Laura, see if there is any additional information to provide back to the committee, and just have you review it offline. You don't have enough information in front of you at the moment to really make this assessment.

DR. MORRISON: June, and then Doug.

DR. LUNNEY: Just to clarify, validity for this purpose, we are looking at whether or not the captured screening response reflects the truth about what is in the record? What are we trying to prove?

DR. WHITE: Yes, that's right. Just to follow up on that, you can have -- From the standpoint of validity, you can have an assessment of face validity. Does the same measure -- what it appears to need to be measuring? That could be via expert judgment, or there could be the kind of question that June is talking about.

So I would wonder if you can have expert judgment of face validity as the relevant validity testing?

DR. MORRISON: Solomon?

DR. LIAO: So again going back to the issue of having consistent standard, we have taken Laura's word on other measures and not asked for her to have to provide the numbers before we voted. So --

DR. BURSTIN: She actually provided the numbers in the form she submitted. The difference here is she didn't submit this measure.

DR. LIAO: Exactly. So I guess I am asking, do we need to table this or is this something that we can take a preliminary vote on today, or go back and revote this issue?

DR. MORRISON: I think we need to confirm it, yes. We can do that. Tracy?

MS. TECCA: Can I just ask a quick question? If we go forward, and I think it would be terrific if the recommendation was that Laura and I could sort of see how we can present the data that is available in a clearer form, but I would say I would love to hear a sort of resolution to this question about what validity -- what do you want to see with regard to the validity, just in terms of the data that would be helpful?

I don't know whether we have it in our database or the combination, but it would be helpful to know the question for sure that we are trying to answer.

DR. BURSTIN: I think we can work offline to provide that to you, but I think a systematic assessment of face validity is one element to get to at least moderate validity. That is why I think more information here would certainly get you closer.

MS. TECCA: I appreciate that, yes.

DR. MORRISON: Last question, Naomi.

MS. NAIERMAN: Just a quick question. Would it help to have any perspective from the AIM folks? Do they have any data that -- because they selected this measure. Just wondering. There may be another source of information.

DR. MORRISON: Eduardo?

DR. BRUERA: Just a brief comment about the issue of the validity and reliability, because in experimental conditions one can find things, and this is going to be proposed for a quality measure for clinical teams to work on, not for doing it under the conditions of having a research person or somebody who is particularly trained to it.

So I think, if the body of knowledge includes how these measures perform in the real world when done by real clinicians, that would be the important part of the information, not whatever you did to publish a paper, and we found with a JCAHO pain fiasco that people can say anything they want. When you tell them to say from zero to 10, they will adhere to any regulatory measure.

Eighty percent of what I dictate in every single medical chart is only for the purpose of regulatory and billing issues, and 20 percent of that has any relationship with patient care.

So we need to be aware that we don't create paper tigers and that we ask for the reliability and validity that, hopefully, will be applicable to the real world clinical scenario as much as possible.

Otherwise, we might run consistently, not only when we talk about this measure, but with any other measures of so called quality, into backlashes of not really improving outcomes.

DR. MORRISON: Tracy, I know you got your thoughts together.

DR. SCHROEPFER: We have had others where we have had the face validity at least, and this one does talk about that it has face validity.

I understand what people are saying, but it just seems like validity -- I mean, in this measure it is pretty much what some of our other measures have been. It is not like does it measure what it purports to measure. It is a conversation was documented or not, and how do you ever test for that.

It doesn't seem like validity really applies so much with some of the more rigorous tests. So this one says they conducted face validity.

DR. MORRISON: Heidi, and then Doug.

MS. BOSSLEY: The one thing with the face validity, though, I would point out, part of the criteria says it must be through a systematic process. So they may have used a systematic process, but it is not yet documented in here. So that is something that, I think, we need to go back and work with them on, and see. But just stating that it has face validity isn't enough to be able to get even a low rating.

DR. MORRISON: Doug.

DR. WHITE: Just a measurement issue, the difference between face validity and criterion validity. We can all sit around here and look at it and say it looks good; it has face validity. But in the real world, the other kind of validity is the criterion validity.

Does the thing really measure what it intends to measure compared to a gold standard would require a chaplain or someone who is really skilled in eliciting spiritual needs separately doing that, going in there, doing that, getting a sense of does this patient really have any spiritual issues to be addressed, and checking one or zero, yes or no, and then comparing that to what you got on your little screen.

I am not saying that we need to do that. I am just sort of pointing out the difference between face and criterion validity.

DR. SCHROEPFER: And I agree with you, and I understand that. I guess I am thinking about our discussions yesterday where, truly, we didn't hold that standard to some of the measures yesterday. But, no, I agree with you.

DR. WHITE: It is tricky when you hear someone say the measure is valid. It is such a thicket.

DR. MORRISON: All right, guys. I

think I am going to close discussion. I want to just -- There are a couple of -- I think there are a couple of issues that I want to summarize on, and then I think a couple of points that I want to sort of just clarify before we move forward, if that is okay, and I am trying to frame this so it will help in terms of the -- because I think we are going to move forward with voting or revoting.

So just a couple of things that I have heard that would help, moving forward. I think what I am hearing around the room is everybody agrees that spirituality is a core component, and I think that the importance of that is -- and I think that the committee is struggling hard with the idea that this is a core component that we think needs to be included.

I think that there are some struggles and concerns about how the measure is being put forward and some of the data behind it. I think the relevance to yesterday

-- people talk about, well, we didn't hold some of the pain standards, for example, to that standard, but we have 20 years of data on pain, and volumes and volumes and volumes of it, how do you assess pain, what are the right instruments, how are the questions asked.

I think that the evidence base for spirituality is lacking. So I would not make the link particularly to pain issues.

I think some of the issues that have been raised, and Laura raised, I think, for the committee are important ones. Everybody understands the importance of spirituality.

I think the problem is that, unlike pain, again to use it, we don't have clear links between the process of assessing spirituality and then a good outcome that addresses spiritual distress.

Even something as simple -- you know, knowing this would be as a chaplaincy consultation, we don't have outcomes data to suggest what percentage of people who identified spiritual distress are alleviated or absolutely alleviated by a chaplaincy visit, and we do have that data for pain.

We do know that, if you identify pain in scale and you give opioids, particularly if you have cancer pain, it works. So I think that is some of the issues that people are struggling with.

I think some of the scientific validity issues that Doug pointed out are very, very critical, that it really -- There is a difference between face validity and asking somebody do you have spiritual distress, and understanding what that means.

As Doug said, there are ways of getting at validity questions. You know, the gold standard may be a chaplaincy consultation, but there are other more comprehensive assessments that get at spiritual distress in a more comprehensive way, and so that type of scientific criterion validity that Doug presented, I think, are very important in terms of making the decision.

I think what you have heard from Laura, which are going forward, is that there is good reliability data, that people answer the same way, that if you ask them this question, they will answer the same way. That is different from what does it mean by having that answer, which comes back to the criterion validity. I mean, what does it mean when they say yes.

So what I would propose is we go back and we revote on the reliability and validity data. I think that we will assume that Laura, because I trust her, has good reliability data. I think for the committee, for the purpose of the vote, I think you can assume that criterion validity is not present, and I think you have to vote your conscience on the face validity piece.

Then what I would say is,

depending upon what we come back when we revote based upon those specifics, then we decide whether to move forward based upon the voting, and I hear people's distress about this one. I do think that this is one that people are -- and I've got -- June is distressed. I see distress.

DR. LUNNEY: I want to clarify. This measure does not measure distress. It measures whether or not a conversation happened.

DR. MORRISON: No, I understand that. I understand that, June. I understand that, but I think what you are hearing is --

DR. LUNNEY: Then criterion to distress assessment is inappropriate.

DR. FINE: I would like to ask the question maybe of Doug. What would the difference in a validity assessment be between -- I see this as a screening question: Was there a discussion or not? Nothing about the quality of discussion, the meaning of discussion, the outcome of the discussion. I read it as a screening question.

What would validity testing look like for a screening question versus an assessment question? I am just trying to understand this. I do not come with any background in statistics.

DR. WHITE: I think June's point is a very important clarification. But it does -- Then it changes the importance of the measure, I think.

If you are asking was any sort of thing about spirituality questioned, that is very different than was that a reliable and valid screening tool for is there a spiritual issue.

If we are not asking the latter -- and that is the point you are making -- then I wonder about what is this thing? What is the utility of the thing? We are not really trying to sort of identify is there a real problem for patients that needs to be followed up on, i.e., is this a real screening tool for a clinical issue.

If that is not what we are measuring, then I guess I need to sort of rethink how I had judged this.

DR. MORRISON: Naomi, and then Helen.

MS. KARP: It seems to me that, if we have this much confusion about what are we even looking at for validity, is this screening, what exactly does screening mean, then there is something wrong with the definition of the way this measure was framed. It is too ambiguous. The wording is too ambiguous. Even if they put the word discussion in there, maybe that is too ambiguous.

So what would be the process for dealing with that? Do we have to reject the whole measure or is there a way to give somebody an opportunity to tweak the wording and then have us look at it again?

DR. MORRISON: Helen, and then Kate.

MS. MARTEL: Could someone from NQF answer that as a process question?

DR. MORRISON: Yes.

DR. BURSTIN: We have before asked a developer to go back and provide further descriptions and refine the numerator statement so that it is clearer, and include definitions, anything like that. You can request that, and I think that is actually very reasonable here to provide more guidance as to what you would expect to be documented to be able to meet the measure. That is perfectly appropriate.

DR. MORRISON: Helen?

MS. MARTEL: So I guess my question would be, to follow on both of those, is just that the clarification for are we asking was the discussion raised? Was the whole issue just raised in a discussion, versus was there, as Doug said, a further --

you know, more depth to that screening than did you even raise the topic of spiritual dimension to a patient, or not.

DR. MORRISON: Kate?

MS. O'MALLEY: My question would be to Helen. Since this question is in the field and is used extensively in the Medicare conditions of participation, when you ask this question, what happens? Are we just asking the question out in the wilderness, and the tree falls, and nobody hears it, or does something happen in relationship to this that needs to be part of the process of considering the value of this measure, this question?

DR. MORRISON: Helen, you want to talk about it?

MS. MARTEL: Well, I guess that is what I am asking the developer or the folks who are promoting the measure, is exactly that. It is to find that out. Is it just raising the discussion, just raising the issue, period, or is it then is there an assumption that then you are doing something about it? You are following through, and then do an assessment, or something, or are you just simply saying, you know, we are raising the topic to give the opportunity for a patient and family to express that they have a spiritual concern, and the next step would be further screening and assessment of something?

MS. TECCA: Let me try to answer that, to a certain extent. I think all of us that are trying to measure hospice, look broadly at the different kinds of things we can measure in palliative care and hospice, we are looking at, of course, not just screening, but then what do you about getting to the outcomes.

I think, if we look across the various dimensions, there are lots of things we don't know exactly why all the different interventions that are gong on work, if they work. So we acknowledge that.

Ideally, you would like to have: Did you have a discussion? What exactly did you ask? What did you do? You would like to have all those pieces. So I think we acknowledge that.

We started, however, this particular measure because, if we go back to the disparity data, which we haven't talked about, but in fact this simple question of did you document that you had the discussion, at the median it is three-quarters of agencies say yes to it, and 25 percent of them only half the time are they documenting a required discussion, in something that everyone here says this is critical that we have this discussion with our patients, everyone around the table.

In the disparity we are noting that, even on this first question, simple screening question, only half of the people are saying yes, 25 percent of the hospices, so out of 100 different hospices that were participating in this, and more, I know in others.

So we don't have that next level. Yes, sure, it would be nice to have the other measures, but this is the first measure, which it seems, based on the fact that we all recognize it is important and we are seeing from the disparity key data that the discussions aren't happening in a lot of hospices, just having a discussion would seem to be important. So that is the -- Does that answer the -- And the question earlier about if asking this question -- Again, the question is did you with that patient have the discussion? That is the question.

DR. MORRISON: All right, Heidi. What do I do?

MS. BOSSLEY: Put me on the spot.

DR. MORRISON: Of course.

MS. BOSSLEY: One thing we could do -- So let me throw out a couple of options. One thing we could do is, based on this conversation, to me, either you can wait and see if additional information from Laura would be helpful.

The other thing to do is revote, based on the conversation you have had, looking at the information again. You could do that, and then we could see where the votes come out. Those, to me, would be two options.

DR. MORRISON: Strong feelings as to should we revote? David?

DR. CASARETT: We spent so much time on this discussion, and I confess, I am actually less clear about this measure than I was an hour ago. So although I was happy to vote the first time, I think we need more clarity and more information. I wouldn't feel comfortable voting at this point.

DR. BURSTIN: We will take it back.

DR. MORRISON: We will take it back.

DR. BURSTIN: We will sit down,

and we have got the questions. We will get the answers and get back to you.

DR. MORRISON: Does that feel good with people? Good. Okay, onward and upward. Great discussion, guys.

I had a feeling about that one. I was hoping to go through the first two first, so that we would be feeling really good before we got it.

So I have got -- Is anybody from RAND on the phone? It is ungodly early on the West Coast. Anthony, do I have anybody from RAND, either Neil, Carol? Neil is here? Fantastic.

We are going to do measure 1626, which is patients admitted to an intensive care unit who have care preferences documented. Would you like to give us a little introduction to the measure? DR. WENGER: Sure. So this measure I developed using the RAND-UCLA methodology that we discussed yesterday. This one is focused upon preferences, and it is whether patients who have been -- vulnerable elder patients who were admitted to the intensive care unit have documentation concerning care preferences within 28 hours of ICU admission.

Documentation of care preferences includes any of the following, including code status: Preference for general aggressiveness of care, mechanical ventilation, any sorts of life sustaining treatment, including dialysis, transfusion, feeding tube, a permanent feeding tube, or that there has been documentation of a care preference discussion, or that such a discussion was attempted or why it couldn't be carried out.

The denominator is simply all vulnerable elders who are admitted to an ICU, and who survived for 48 hours.

This measure has been implemented a variety of times now, and there is a clear performance gap in the three different studies, some with very tiny ends. It is satisfied less than a fifth of the time. In the larger study it was satisfied 46 percent of the time. That was published last year in a sample of 369 patients.

It has excellent reliability. Its validity derives from the process with which it was developed, which is both literature based and expert input, considering the strength of the process/outcome link.

It also is part of a group of measures that has been demonstrated to be associated with better survival, which one wouldn't expect this measure necessarily alone to be associated with, but also better functional outcome among vulnerable older individuals.

We are not aware of any measures with which it would need to be harmonized.

DR. MORRISON: Fantastic, Neil. Thank you very much. Tracy, I think I have you as the lead discussant on this. Right?

DR. SCHROEPFER: Yes. So he just

described the numerator and the denominator. So I am just going to go into the summary of the reviewers and how they felt, and I will skip the high impact.

The opportunity for improvement, the demonstrated performance gap: Pretty much it was all highs except for one moderate and a couple of insufficients. People felt this performance gap was well documented. There was a reasonable discussion of it.

The only question was that they felt that the issue was important to all ICU patients, not just the vulnerable. So they questioned that.

In terms of the evidence, so here in terms of the quantity, it was high to moderate, and there was just a few in terms of insufficient and low.

Quality, again, mostly moderate, and consistency high to moderate. On the positive side, felt studies demonstrated all three, quantity, quality and consistency.

Even if there weren't clinical trials, preferences are an important step in care planning, and there was a code expert panel, and then evidence.

On the other side, they felt like there was only ungraded expert panel recommendations. They felt the population and preferences were vague, and that there was a lack of information.

In terms of reliability, it was high again, to moderate, but there were actually three lows. They felt that the kappa -- that is the inter-rater reliability -- that was good, the measure well defined. The numerator was reliable, but that the measure itself was not directly tested, and this came up several times, that it was tested as part of a group of other measures.

Also, it was not clear to one individual how to define care preferences, because code status doesn't always match care preferences.

In terms of validity, again expert panel validity through a code, a code 3 and assist, and so that was satisfactory, but some people felt that the face validity was a low level test. Again, items were not tested directly, but part of a larger tool, and it was hard to -- it would be hard to measure whether, if there was a failure to attempt to elicit the patient preferences, that would be unknown, and that was very significant.

In terms of usability, this was moderate. It has been used by the UCLA Medical Center, and followed, but there really wasn't enough information to determine about the usefulness for public reporting. It is just not clear how they would report, the usefulness for quality improvement.

There was a feeling that it is not clear how useful it would be, because of the inconsistency in recording and documenting whether there was a discussion. So there may be a presence sometimes as an advance directive, but in terms of this discussion being recorded, there was concern, and they felt that in that respect, it would not be useful.

Feasibility: Feasibility, for the most part, was moderate, and the feelings there were that documentation of patient preferences -- it is not consistent, and maybe this would push for consistency, and one person said, but it is being used at UCLA. So that shows that it is feasible, but again there is this concern about could you actually capture this using electronic data capture. Unless you did record progress notes electronically, really, the data collection would be a manual process, and that is going to leave it open to errors and problems.

In terms of suitability, there were four yeses and three no. Even in the yeses -- So the yeses talked about that very much it impacts a significant number of patients, and knowing preferences can impact the treatment decisions.

They felt that there was ample documentation for implementing the measure, and it could lead to improvement and better standardization of documentation. But even the yeses had some concerns, and that is that the measure requires chart audits and some documentation in paper charts. So again the data collection could be burdensome.

The documenting professional discipline was not specified. Physicians and families -- so this is a comment, I think, more than anything -- often know patient preferences, and ignore them.

For the noes, more clarify and evidence needed. Numerator: You can't count, again, on the data being in the chart, being reliable. So the numerator is not reliable, and they felt that the reliability and validity testing no strong. Then the definition of care preferences was just too broad to be meaningful.

They felt like in states where there is a MOLST and POLST forms, it would be easier to gather that information.

So that is pretty much it.

DR. MORRISON: thanks, Naomi. Open for discussion. Russ, Rick. Russ?

DR. ACEVEDO: A lot of those negative comments came from me. So I might as well start. The determination of the patient wishes and plans is crucial. So I want to make sure that my other statements -- I am just not sure this measure measures it, in that one of the things that jumped out at me is the code status, that if it is documented, already your numerator, you have satisfied the requirement, or it may have no relationship to what the care plan for that patient is.

So that I immediately found as -- because my residents always tell me, oh, yes, I talked to the family, and the patient is at full code. Well, that still doesn't tell me what the patient's goals of care are, what do they want to accomplish, what things are doable, not doable.

It is a long, meaningful discussion that needs to occur, very important to occur, but specifically the way this is written, my residents had the discussion, yes, we want resuscitation, I have met the requirements of this measure.

So that is written too broad to be meaningful. We are in a state that has a MOLST form, which is very helpful when it is filled out. So for instance, if this measure were to go forward, then I think we will see more use of the MOLST form.

Sorry, for those who don't know, Medical Orders for Life Sustaining Treatment. So it is a form where you essentially -- they standardize the process of going over treatment decisions. We talk about resuscitation, ventilation, but you also talk about artificial feeding, the whole host of treatment decisions, and we have documentation.

I think something like that -- If this were to promote that, I would say yes, but I think the way this is written now, I am not sure I am going to capture the information.

One last thing, and I will stop. When I am having a discussion -- So a patient gets admitted to my unit, and I am having a discussion with either the patient or the surrogate. We are discussing what our goals of care are, and we are implementing it.

What this is going to -- So I have had that discussion that is based on the admitting plans, but there is not a specific note there that says -- I mean, the note would say that we have discussed the admitting plan. I am not sure then that then satisfies that I have specifically written down the patient's preferences.

DR. MORRISON: Rick, and then Naomi.

DR. GOLDSTEIN: I mostly have a question, and it has to do with the usability of this, this measure as it is prescribed. So given the fact that a lot of these patient for the first 48 hours in the ICU are not going to be communicative -- I understand why it is important to have on record even prior to their transfer to the ICU that these conversations have been held or that their preferences are known, but what is the strongest argument for this measure being designed this particular way?

Is this really just an ICU document, a documentation issue?

DR. MORRISON: Neil, can you answer that, and can you also speak a little bit to the issue that was raised before, because I know you have provided some journal publications supporting the feasibility, reliability and validity of the measure, but I think it might be helpful for the committee to hear a little bit more what is in those publications.

DR. WENGER: Right. I am not so sure that I could hear the entire most recent question. Can you paraphrase it?

DR. MORRISON: Let me try. Go ahead, Rick.

DR. GOLDSTEIN: It is really a usability issue for the measure as it is written. So given the status of the majority of these patients or many of these patients in this first 48 hours in the ICU, why is this measure as applied in this time frame important?

Is it an ICU documentation need or, really, shouldn't the question be asked about the preferences for these patients even prior to the transfer in the ICU?

DR. WENGER: There is no question that, from a theoretical perspective, advance care planning should be advanced. This measure, like all of our quality measures, attempts to hit at a very low bar, and when it, therefore, is failed, it demonstrates clear areas where improved care is needed.

The low bar is felt that, if someone decompensates far enough and they are a vulnerable individual receiving hospital care, that for sure there ought to be documentation concerning preferences within two days in the intensive care unit.

That in no way discounts the fact that there should have been documentation in the outpatient setting before they ever got admitted and early in hospitalization soon after admission, but if you are trying to identify a group of people where our expert panels felt that there was no question that there should be documentation concerning their preferences to guide their care, it would be within the first couple of days of intensive care. So it is not a regulatory issue. It is based on care needs.

To attempt to address the usability issue a bit more: So only in one of the three applications of this, albeit that the other two were small, were UCLA records used.

ACOVE-1 was from two large managed care insurers, one in the northeast and one in the southwest from a whole variety of different hospitals, and the ASSIST measures were also only partially measured at UCLA, but also in other venues.

So this has been used in a variety of different hospitals and different ICUs, and there has been very little difficulty in identifying documentation concerning preferences, and even though there is a concern that someone may simply jot down something that doesn't reflect in depth conversations, it occurs less than half the time for this group.

DR. MORRISON: Naomi.

MS. KARP: I guess I would make the point in response to that, that it doesn't say that it requires a discussion of preferences with the patient. I agree with you that many, if not most, of the patients couldn't have it, but it could come from an advance directive. It could come from a conversation with a legally authorized proxy, a default surrogate.

So there is a potential to get the information from a lot of different sources. Also, I guess there was some discussion of MOLST and POLST, but I just wanted to kind of emphasizes that, to the degree -- You know, it is a paradigm that is spreading, and 12 states -- Kate, correct me if I am wrong. I think about 12 states are endorsed POLST states, and a lot of other states are implementing it.

To the extent that this measure could really encourage the spread of POLST, I think it is really important, because POLST is a form that is very specific. It covers most of these measures.

I agree, it is a problem if we are just going to have one of these particular kind of interventions documented, because that is not going to get us that far, but if you use a POLST or a MOLST or a POST or whatever the state calls it as a form, you are going to get it very comprehensively, in a defined way that is going to be on a bright pink or yellow piece of paper in the record or, hopefully, one of these days electronically.

So I would really think this could get us a long way toward the spread of POLST.

DR. MORRISON: Thanks, Naomi. I have got Doug, and then I have got Kate.

DR. WHITE: I would just echo what Naomi said. Speaking from the perspective of an intensivist, this is a huge, huge problem. The care that is delivered in ICU is often un-patient centered with no documentation of anything and no patient preferences, no conversations with surrogates in those early days.

Russell, I think your group does better than most groups, even in just that, yes, we got the code status. I certainly wish the measure held people to a higher standard, but thinking about where we are in the process of nudging things along, I think this is a pretty acceptable place to start, and relative to a lot of other measures we have looked at related to sort of what we do in ICUs, it is unobjectionable, ethically.

There is just -- All the ethical vectors point in the direction of you should at least be doing that. It doesn't have anything to do with what your outcomes are. It is just this is a process measures; you know, you should do this.

We've got lots of pretty good data about reliability and validity. I mean, it seems to me a little bit milk toast, but a good measure, nonetheless, with enough supporting scientific evidence that I look at it favorably.

DR. MORRISON: Kate.

MS. O'MALLEY: A question for

Neil. I am just curious about hearing that meeting the bar for this would be met, even if you had code status in ICU patients. Since this was a bundled measure when it was tested, I am wondering if you know the degree to which patients in ICUs don't have even a code status determined for them in that setting.

DR. WENGER: More than half the time.

DR. MORRISON: Neil, you are not here, but there are a lot of people in sort of stunned shock, and you know, I will tell you that, speaking as a physician, it is astounding. But, yes, that is what we see, and it is stunning.

DR. WENGER: It is presumed.

DR. MORRISON: Other questions for Neil? People comfortable moving to voting? Again, really good discussion. So let's move to voting.

All right, guys. Importance of the measure: Performance gap, does it meet a performance gap?

MR. COLCHAMIRO: Nineteen high, one moderate, zero low, zero insufficient evidence.

DR. MORRISON: Evidence or outcome?

MR. COLCHAMIRO: Seven yes, 13 no.

MS. HILL: I think that people aren't understanding that question.

DR. MORRISON: We are just going to skip it.

DR. WHITE: It is a teachable moment.

DR. MORRISON: You know what, Doug? I will come back to the teachable moment afterward. You are exactly right, but I figured, since Helen and Heidi were not giving me grief about it, I was just going to let it fly. All right, we will come back to the teachable moment.

Quantity of studies and body of evidence? We will come back to the teachable moment.

MR. COLCHAMIRO: Eleven high, nine moderate, zero low, zero insufficient evidence.

DR. MORRISON: Quality of the body of evidence?

MR. COLCHAMIRO: Twelve high, eight moderate, zero low, zero insufficient evidence.

DR. MORRISON: Consistency?

MR. COLCHAMIRO: Fifteen high, five moderate, zero low, zero insufficient evidence.

DR. MORRISON: Reliability?

MR. COLCHAMIRO: Ten high, nine moderate, one low, zero insufficient evidence.

DR. MORRISON: Validity?

MR. COLCHAMIRO: Seven high, 12 moderate, one low, zero insufficient evidence.

DR. MORRISON: Disparities?

MR. COLCHAMIRO: Nine high, seven moderate, seven low, four insufficient

evidence.

DR. MORRISON: Usability?

MR. COLCHAMIRO: Ten high, eight moderate, one low, one insufficient evidence.

DR. MORRISON: Feasibility?

MR. COLCHAMIRO: Seven high, eight moderate, four low, one insufficient evidence.

DR. MORRISON: And overall for endorsement? All right, would the uncertain person out there please vote.

MR. COLCHAMIRO: Twenty yes, zero no, zero abstentions.

DR. MORRISON: Terrific. Neil, I think we have you for one more today. Is that right?

DR. WENGER: Correct.

DR. MORRISON: Thanks very much. As you know, it is endorsed, 20 people endorsed, and I just wanted to thank you, and would you just pass on your thanks to your group who has done a really tremendous amount of work to push this work forward.

DR. WENGER: Appreciate that.

DR. MORRISON: Okay, Doug. You have your teachable moment.

DR. WHITE: Right. It's gone now. What was it? What was the issue?

DR. MORRISON: What is an outcome measure, sir? You have your teachable moment.

DR. WHITE: Right. A process of feedback from on high, because I have a very, very short attention span.

The issue we have all been voting on, is this thing a health outcome or is this a process measure along the way to a health outcome? It seems like a lot of these things where you are going to the chart and seeing, there is a certain process of care accomplished. That is not a health outcome. That is a process of care.

DR. CASARETT: That is what people have been voting on? I'm not sure how people have been voting on that. Is that what people have been doing, seeing these as outcome measures?

DR. MORRISON: The question is, is this an outcome measure? It is a straight yes or no.

MS. BOSSLEY: We know that that one needs a little work. We have been trying to figure out what to do. We don't know what to do. So I think, for the purposes of your discussion today, we are probably going to -- You could skip over that one, and we are probably going to throw out the responses on that, because at this point I am not sure of that.

We could even explain what you would have been voting on and how you voted. So I think we are going to --

DR. WHITE: So you are saying you guys thought the question had face validity, and in practice it turned out to not work?

MS. BOSSLEY: That is it. That's it.

DR. MORRISON: Naomi?

MS. KARP: Just to clarify the process, isn't that supposed to be a threshold question, and it determines whether you then go on to the next three?

DR. MORRISON: No. No.

DR. BURSTIN: Actually, just two seconds. The logic of that question is that, if it is an outcome, there is a bit of a pass on evidence, particularly for adverse outcomes, but for all -- Most of the measures we've been talking about have been process measures. So it is not a threshold, and you move on and discuss the quality, consistency anyway. Yes.

DR. MORRISON: Okay. The last measure before our break: Dr. Hanson, we have hospice and palliative care treatment preferences. We are on a treatment preferences roll. This comes from UNC Chapel Hill group. Laura, are you with us on the phone still, hanging in there?

DR. HANSON: Yes, I am.

DR. MORRISON: Would you like to present this measure?

DR. HANSON: I can tell you briefly about it. It is developed through the same process as the other measures you have heard from our group, a two-step process with initial development and testing in a hospice population, and then extension into a hospital based, seriously ill population with hospital based palliative care services.

This measure I really appreciated, although -- I really appreciated the discussion of the RAND measure, because I think a lot of that discussion helped frame the background for this quality measure as well.

I would say that the operational definition of the two quality measures is quite comparable, but the denominator population is different. The denominator population for this quality measure is the same as you have heard from our group before, hospice and seriously ill hospitalized patients with palliative care.

We have those same exclusions with palliative care for at least one day, and hospice enrollment for at least seven days. The description of the numerator -- so the documentation that is required for evidence that there has been a discussion of life sustaining treatment preferences -- includes the presence of documentation of discussion of those preferences with the patient or, if the patient is not capable, with their designated surrogate or review of advance directive, and salient to, I think, some of the earlier discussion, we have a specific description in our operational description that a brief statement that the patient is full code, for example, is not accepted as evidence that a discussion has occurred, or a brief statement that the patient is do not resuscitate is not accepted as evidence that their preferences have been discussed.

This received a high rating from our technical expert panel with a rating of 4.04 out of 5.00 possible points. Inter-rater reliability was excellent, a kappa of one, and we have evidence for a gap in performance. In the hospice sample, this was documented for 82 percent of patients, and in the seriously ill hospitalized sample for an overall level of 67 percent, and that overall level was differential, depending on whether specialty palliative care was involved. So when specialty palliative care was involved, the treatment preferences with goals of care and evidence for discussion was documented 91 percent of the time.

Fantastic. Thank you so much, Laura. Just to sort of frame this -- and, Laura, correct me if I am wrong; I know you were here for the prior discussion -- this is a very similar measure to what the RAND group proposed. It is a little more comprehensive, as you said, that code status is not acceptable, simple documentation. The major difference is the denominator population is different, and the setting is different.

As you pointed out, the bar may be different in hospice and palliative care than it currently is in the intensive care unit. Actually, I think Doug and Neil pointed that out.

So I have Bob as the leader for the discussion, and then I will open it up to comments from the group. Bob?

DR. FINE: This one was endorsed by six of the nine people who reviewed it, and the three people who said, no, we have reservations, were concerned about chart abstraction, kind of the ability to carry this out across different settings by different nurses. I think the steward, though, has kind of answered that question, if I remember, inter-rater reliability.

One of you all expressed concerns about what was really included in the numerator and the denominator exclusion of less than seven days on hospice. We have talked about that before. Overall, it got high marks from people.

As a parenthetical aside and related to some of our earlier discussion, I think that it is interesting that their benefit statement briefly explains the benefits, improvements in quality envisioned by use of this measure. Quote, "Use of the treatment preference's quality measure will improve attention to this important practice," which is kind of one of my whole points about trying to just assess spirituality. Forget about the quality of the assessment; just draw attention to the -- I just thought it was interesting. Couldn't resist putting that in. So now I am just drawing attention to things.

I've got no other comments. It was basically favorably reviewed, with a couple of exceptions that just had some worries about carrying it out.

DR. MORRISON: Thanks, Bob. Appreciate the opinion. I've got Helen and then Naomi, and then Doug.

MS. MARTEL: So my question is about the denominator being patients in specialty care in the acute care hospital. Our inpatient teams follow patients from the acute care into ICU, and they see them in both places and consecutively.

So, to me, I was presuming there was an overlap between this measure and the last one. Is that not true?

DR. MORRISON: Let me try and answer that, and I will turn to Heidi as well. I think part of the issue, Helen, is, yes, there is overlap. I think part of the problem is that there are people there -- In terms of the NQF endorsement process, the North Carolina measures weren't tested in the ICU population, and the ACOVE measure wasn't tested outside of the ICU.

So you are right. You have a

potential overlapping Venn diagrams, but not -- overlapping, but not the same, because are were people cared for within palliative care teams who are not in the ICU, and vice versa.

I'm sorry. Naomi, where did you go? You are down. And David -- or, Doug, you are up.

DR. WHITE: Laura, thanks. Again, this seems like a great measure with lots of face validity and scientific acceptability -- scientific testing and acceptability. So mine is just a very small question, and I think the inpatient palliative care consultation is where the palliative care service is consulted, because of difficulties around goals of care, and what is documented as -- this is sort of for the vitalest patient. This patient is full code.

It seems like that is a very clear statement of preference that, just documented in that way, is an appropriate documentation, albeit controversial regarding the patient's goals, but wouldn't fulfill your criteria for having had a -- appropriately satisfied the numerator conditions. Can you help us understand that a little?

DR. HANSON: Because of the focus of the quality measure on documenting patient preferences, we required some evidence that those preferences were brought forward, either through an advance directive or through a direct discussion with the patient or with the patient's surrogate.

If the only documented statement was this patient is full code, then we did not accept that in the numerator. If the documentation indicated "discussed with patient, and patient is full code," we accepted that in the numerator, and it was really because of the need to make clear that the goal is to document communication of preferences or respect for preferences rather than to write an order or make a medical decision.

DR. WHITE: Thanks.

DR. MORRISON: Rick?

DR. GOLDSTEIN: I had had some questions about inclusion criteria, and the question that I -- The main sticking point I have with this is that, if you wait -- If you remove all the patients who have died in the first seven days in hospice without a conversation, you have lot 30 percent of the patients that are referred to hospice. I just wonder if that is too loose of a criteria.

Then in a minor way, in terms of the inpatient palliative care programs, I think, as certainly in pediatric palliative care but also as programs mature, I think having the DNR conversation at the first meeting is not really what we do a lot of the time.

A lot of times, we are really there for clarification of goals, and if we can have that conversation or not has sort of more to do with whether we have made a full assessment of where the family is or where the patient is and how available they are to that conversation.

So I am not so sure that its use for palliative care in the first 24 hours is really -- it has the same validity as we are sort of imaging it to have.

DR. MORRISON: Solomon.

DR. LIAO: So mine is a follow-up question to Doug and also to the reviewer's concerns about feasibility. So, Laura, could you help us to understand a little bit more about the feasibility question of manually extracting from the chart the distinction between a simple full code order and a full discussion about code status?

DR. HANSON: The basic operational definition is pretty much as you see it in the numerator details that are there, and our nurse abstracters did not find any difficulty and really had this very strong kappa of one inter-rater reliability.

The criteria were really to look for evidence that there was a statement like "discussed," a statement of communication or evidence that it was grounded in a written advance directive that the provider had reviewed.

I guess I don't know how to say more than that. They did not find it difficult from a feasibility standpoint beyond the usual feasibility concerns of a requirement for medical record review which, once you put that in place, does ask for a little more time commitment. But actually abstracting the information, they did not find to be difficult.

Remember, this is a two-step process. So the inter-rater reliability was done with nurse abstracters in a hospital setting, but we also had a multitude of nurse abstracters working in 22 different hospices with different documentation methods, and they did not find feasibility challenges with this particular quality measure.

I want to go back to the comment right before yours. The denominator excludes palliative care patients who are seeing a palliative care provider for less than one day, but this measure is not time dependent. it does not say that the documentation has to be recorded within the first 24 hours, just that it has to be recorded during the time that palliative care is being delivered.

DR. MORRISON: Thanks for the clarification, Laura. David, and then Rick.

DR. CASARETT: Hey, Laura, it is Dave Casarett. Just a quick question about reliability.

I notice that you had the kappa of one, which is great, something to which all of us aspire to, but if I am reading this right, it is a kappa of one with two nurses in the non-palliative care sample of 20 patients.

DR. HANSON: Correct. Small sample.

DR. CASARETT: It seems like there is more there. Do you have any sense, either formally or informally, of more real world kappas, either in a larger sample with other nurses or things you have heard from hospices, for instance, who have tried to implement this, just to give us a sense of what it might look like in the real world?

DR. HANSON: I wish I did, David. I really can't answer that. That is the extent of the inter-rater reliability data that we have, and when the kappa was so good with 20, we felt very comfortable with that, but we don't have additional inter-rater reliability, and I am not confident that other people do, other than what you have heard from Neil's measure right before this.

DR. MORRISON: Are folks comfortable moving to a vote? Yes? Okay. Lindsey has got it up already.

Performance gap?

MR. COLCHAMIRO: Sixteen high,

three moderate, one low, zero insufficient evidence.

DR. MORRISON: Quantity of studies in the body of evidence?

MR. COLCHAMIRO: Sixteen high, four moderate, zero low, zero insufficient evidence.

DR. MORRISON: Quality?

MR. COLCHAMIRO: Thirteen high, seven moderate, zero low, and zero insufficient evidence.

DR. MORRISON: Consistency of results?

MR. COLCHAMIRO: Nineteen high, one moderate, zero low, zero insufficient evidence.

DR. MORRISON: Scientific acceptability.

MR. COLCHAMIRO: Twelve high, eight moderate, zero low, zero insufficient evidence.

DR. MORRISON: Validity?

MR. COLCHAMIRO: Twelve high, seven moderate, one low, zero insufficient evidence.

DR. MORRISON: Disparities? I need everybody to click one more time. Thanks.

MR. COLCHAMIRO: Eight high, six moderate, one low, five insufficient evidence.

DR. MORRISON: Usability?

MR. COLCHAMIRO: Thirteen high, five moderate, one low, one insufficient evidence.

DR. MORRISON: Feasibility?

MR. COLCHAMIRO: Eight high, 10 moderate, two low, zero insufficient evidence.

DR. MORRISON: And overall endorsement?

MR. COLCHAMIRO: Nineteen yes, one no, zero abstentions.

DR. MORRISON: So that brings us to the break. Laura, I am sorry. Heidi has got a --

MS. BOSSLEY: I'm sorry. I have always got to come in. So I think the one question is whether or not we have -- There are two related measures. The question of harmonization, of course, will come up.

The other piece, though, that we will need to bring back to you on another call is there is what I think would be either a related or competing measure within the NQF portfolio now that is endorsed on advance care plans.

So that, actually, is in the process. The developer is updating the form right now, including evidence, etcetera, because we plan on having that go through maintenance in another project. We are looking at whether we should move it into this project and have you look at all three measures on a conference call, and determine what the next steps are, because that one is ages 65, etcetera.

I will tell you now, and I will

say it on the other call, I developed that measure way back when. So I am going to recuse myself from that discussion when that does occur, but that will be just heads up coming to you a conference call soon.

DR. MORRISON: And I think, Heidi, I know that there is also the issue about some harmonization from some of the pain measures yesterday between the UNC group and the RAND group, and I know that there is -- now that those are through, I think we can have that discussion.

Laura, I think you are done with us. Is that correct?

DR. HANSON: I am. I think I am going to go see some palliative care patients.

DR. MORRISON: I wanted to again express the committee's incredible thanks for the work that you did as part of the PEACE Project to get these measures forward to us, and also to really thank you for the quality of the work that you did on the application process. It made it really, really easy to review these. There was a tremendous amount of data that you provided. It was extraordinarily helpful, and thank you very, very much for the work of you and your group in terms of moving this forward, and enjoy the North Carolina day. I hope it is cooler than it is up here in D.C.

DR. HANSON: Well, it is 101. Thank you all very much. I appreciate your attention to this area. I am glad to be part of it.

DR. MORRISON: So, guys, why don't we take a 15-minute break, and reconvene at 10 minutes after the hour, which only puts us 10 minutes behind for the day, and Dr. Lunney is not nearly as long-winded as I am. So she will get us back on track.

(Whereupon, the foregoing matter went off the record at 10:54 a.m. and resumed at 11:14 a.m.)

DR. LUNNEY: All right. As we

reconvene, we are still continuing with the Steering Committee discussion of care preferences. We are now on measurement 0209, comfortable dying.

Do we have a developer to comment on this? Carol, do you want to comment on this?

MS. SPENCE: Yes. This mic is working okay? Great.

I just want to do a couple of things. One is to give you just a little bit of background on the development of this measure. This measure actually has a very long history. It was begun with the Outcomes Forum Task Force, which convened back in 1998. So the hospice community has been thinking about quality and measurement of quality for quite sometime.

That outcomes forum developed actually a set of measures which were based on a document that was called a Pathway for Patient Care at the End of Life.

That document identified three end result outcome measures, one of which was safe and comfortable dying, and then this measure, the specific comfortable dying measure, was then developed by the Outcomes -- one of six that were developed by the Outcomes Forum Task Force. Then we moved ahead with ongoing data collection after that.

So the focus of this measure is that pain is brought to a comfortable level within 48 hours of the initial assessment on admission to hospice. So, obviously, it is looking at pain management at the start of hospice care, which addresses a basic aspect, obviously, of hospice practice. But we also feel that it is useful and meaningful for consumers, providers, and payers as well.

So data collection is done on admission with a follow-up very shortly thereafter, and then NHPCO has been supporting this measure for quite sometime in terms of data submission by participating hospices, and also then national reporting.

So the hospices that choose to -- and this is, again, voluntary data submission -- will implement the measure and then give us data on a quarterly basis. We then run an analysis. They give us the data online. We then run an analysis, and prepare a national level result report, which we post, and then the hospices are able to compare their results to the national results.

So I just want to reiterate, though, that when you look at the disparities piece that this is a voluntary data collection. I would expect to see even greater disparities in performance among hospices when this is implemented universally. These are highly motivated hospices who currently use this measure and give us the data.

So the basic process for the measure is two steps. On admission or, I should say, at the initial assessment, the nurse who is doing that initial physical assessment asks the patient -- this is prior to doing any pain assessment -- are you uncomfortable because of pain.

So for patients who answer yes to that question, they are entered into the measure. The nurse goes on then to do the comprehensive pain assessment as appropriate for that patient, and then begins intervention.

Then within 72 hours of that initial assessment, the patient is contacted and asked was their pain brought to a comfortable level within 48 hours of the start of hospice services, and then that becomes the numerator.

We feel this measure aligns with the HHS National Quality Strategy and that it ensures patient choice for desired level of treatment, because it truly reflects patient goals for pain management.

Also, we feel that that is

probably the primary benefit of this measure, the fact that it does reflect patient goals. It is not setting a particular assessment -- or mandating a particular set of assessment tools to be used. It allows the patient to decide if they are comfortable or not, and recognizes that, for example, on a rating scale a three may not mean the same thing to each patient.

Those rating scales are good for looking at either populations or within patient trending just to evaluate how you are doing with that particular patient, but simply asking the patient if they are comfortable and then allowing them to -- The clinician is then free to both assess with whatever tools are appropriate for that patient, and also then the patient gets to decide if a six constitutes comfort for them, for example.

The other thing I just wanted to mention in terms of the discussion previously about what is an outcome -- This is an outcome measure, and as such, based on conversation with NQF -- yes, with NQF staff -- the emphasis there was for structure and process measures, presentation of evidence was critical, linking it to an outcome; but this was an outcome, the evidence requirements were not as stringent.

So that was my understanding, perhaps a misunderstanding, but that was our understanding. So consequently, we did link it to the -- well, the national -- the palliative framework, the NQF palliative framework, but did not provide a ton of citations or other evidence for it, because it was an outcome measure. But as previously been discussed, there is a great deal of that for pain and pain with hospice also available.

DR. LUNNEY: Thank you, Carol. Sean, am I correct that you will introduce it from our perspective?

DR. MORRISON: It is me. So Carol has done a really nice job of presenting the measure, and I am not going to go over anything she repeated.

I think there was remarkable consistency in terms of the preliminary evaluations from the group. There was -- The reviewers felt that there was a tremendous opportunity for improvement with moderate to high levels.

There was agreement, Carol, that it was a health outcome, that the quantity of the evidence was moderate to high with one exception. The quality of the evidence was rated as moderate to high with one reviewer thinking it was low, and that the consistency of the evidence, again, was moderate to high.

High level of reliability and validity, some split voting in terms of the disparities issues, and then very high endorsement of both the usability and the feasibility issues, I think, largely because of the data that NHPCO provided for that.

All but one person recommended

this for endorsement. I think one of the outside reviewers was concerned about only one source of data, and he or she may or may not want to feel comfortable about talking about that.

I think, from my review perspective, there are just two questions that I have that, I think, would be helpful to highlight, Carol, if you could -- actually, one question, one statement.

So the statement is: This is -- It is a hospice population, I guess. The question, Carol, is that when you have looked at the performance of this measure, it looks like 19 percent, plus or minus, of folks achieve the goal of comfort after 48 hours.

My question to you, because this came up in yesterday's conversation, is: I am presuming that you guys -- that there is not a number that you are targeting, and that this is -- that we don't know what the right number is.

MS. SPENCE: That is correct, Sean, and thank you for bringing that up, because I did mean to make that point. We have not set a benchmark for this. In fact, the discussion -- This was actually back before I was even with NHPCO. I was on the committee that developed this measure.

Our intention was never to expect 100 percent on this. In fact, when we have a hospice coming in whose numbers come out to 100 percent, we look hard at that. We actually in some cases will call them up and check with them about their data.

Again, this was a wide expert group that was together, but it was composed of primarily clinicians who fully understand no one expects that a patient with uncontrolled neuropathic pain is going to become comfortable within 48 hours. So that would be a totally unreasonable expectation.

On the other side, we have not attempted to set a benchmark for this. This has definitely just been for hospices to compare themselves against what other people, other hospices, are doing.

DR. MORRISON: That is it for me, June.

DR. LUNNEY: Do we have any questions from the group? Naomi?

MS. NAIERMAN: I have two questions. Obviously, this is chart based?

MS. SPENCE: It is chart based in that, when the questions are asked, they are recorded. However, there are various ways that it can be extracted. There are a number of software companies who do hospice software who have incorporated the question into their software. So it can be electronically extracted.

There is also a measure sheet, a measure tracking sheet that we provide that hospices can use. So they don't have to make it official part of the record. They can insert that in there and use it, so it is more readily, easily extractable if you do have a paper record.

MS. NAIERMAN: So having seen it as self-reported by the nurses, and they have -- How is the bias, possible bias, controlled here?

MS. SPENCE: Meaning that the nurse -- because it is the nurses -- In the same way as any other documentation. It is open to a certain level of bias. They are supposed to -- Again, the instructions, if they are following, are -- We actually script the question. They are supposed to ask the question as written and specified by the measure prior to beginning an assessment, and that is part of the instructions. They are not to infer or impute a patient's answer to that question.

MS. NAIERMAN: Another question has to do with those that are cognitively impaired. I see that it is a little bit over 18 percent. So let's just round it as to about one-fifth of the patients.

To the extent that hospices are present in nursing homes, I wonder if there has been any thought given to assessing pain in folks with dementia, which would not rely on their self-reporting.

MS. SPENCE: There are some hospices that we know are doing that. That would be for future testing for this, for someone who was totally unable to respond, although you do need to take into consideration various levels of cognitive impairment.

There are patients who cannot respond using a zero to 10 scale, who can tell you yes or no if they are comfortable. Again, that is one of the, I think, advantages of this measure, is that while you do need a patient who can respond to you with a certain level of understanding, it is more inclusive, because it doesn't specify the assessment tools to be used.

DR. LUNNEY: Thank you. Michael?

DR. LEPORE: The current and intended use of the measure is at least partially for quality improvement, and I wonder if you could speak to the fact that over the two years of data being submitted, the percent of patients reporting being uncomfortable due to pain didn't change at all. So to what extent is it being used currently for quality improvement? It certainly seems like it can.

MS. SPENCE: Yes, it can, and again this is totally voluntary. We have got -- Again, I can lay this more at the feet of a marketing or lack thereof than anything else, and it is more of a programmatic issue, I believe, but we have in the last couple of years really started looking very strongly at material development and awareness, and if this does move forward and if it does become a CMS mandated measure, all of that is going to change very drastically.

DR. LUNNEY: Doug?

DR. WHITE: This is sort of a general question about the goals of the NQF as it relates to measures that are not benchmarked.

I remember one of our criteria had to do with both can it be a quality measure and its implications for public reporting. How should we think about that when a measure is not benchmarked?

MS. BOSSLEY: Great question. In general, the criteria and our process does not say that it has to have a benchmark, and for the most part, most measures that we have endorsed don't.

What you looked at yesterday morning did recommend a benchmarking, but again we don't require it. It is not something that needs to be, and I think the expectation from our viewpoint is most measures are not specified to the point where you could actually get 100 percent anyway.

So to us, the expectation is not that that would occur. How others may implement it is something that I think no one knows how it will be, but the way it is being used right now, from our viewpoint and the criteria, is perfectly appropriate.

DR. LUNNEY: David?

DR. CASARETT: Thanks. I had a question about the way that the responses are coded. So pain, obviously, in this setting, is often fluid, good days, bad days, good hours, bad hours, break-through pain and so on. So I am trying to get a sense of how people respond to this question.

So if somebody is asked this question and they had severe pain in the morning, it is okay now -- I'm not sure what the right answer is, but do you have a sense of how people are actually answering this in the setting of pain that may change?

MS. SPENCE: Well, again it is are you uncomfortable, with the idea of then, as the assessment moves forward, do you want us to do something about it, is really, really what that is getting at, is what is that patient -- what are their goals. It really opens the discussion of what are their goals for pain management.

DR. LUNNEY: Martha, we can't -- sorry, no. I'm sorry -- or can we? I am conflicted here. I am being told, yes, we can invite your comment.

MS. BOSSLEY: If it would be helpful to the conversation, I would say go ahead, yes.

DR. LUNNEY: Go ahead, but be helpful.

MS. TECCA: It is my goal. In response to the issue about at this moment in time, are you comfortable, the actual question is to be asked 48 to 72 hours afterward, was your pain brought to a comfortable level within 48 hours. So it is not really like, right now, how do you feel. It is asking them to look at a -- It is that point in time, but it is not that second how do you feel. So it might make that a little bit better.

DR. CASARETT: I guess what I was trying to -- I was trying to figure out how I would answer that question, if I was asked whether my pain was controlled to a comfortable level within 48 hours, if I had had some good hours and bad hours within that time period. Again, I don't think there is a right or wrong answer.

MS. TECCA: Well, again, you get into that when you are doing -- depending on what you are able to do with a patient -- the comprehensive. If you are doing something like the BPI, you are going to have a concept of how the pain fluctuates over time, and when you are looking at character and so forth on an initial assessment, you are going to try to go into that as fully as possible.

That will help elucidate what that patient means by that first answer.

DR. LUNNEY: Are there any other questions from the panel? One more.

MS. KALEN: I know that this measure is designed to be assessed when the patient is first evaluated on admission to hospice, but we hope that people will be there for more than 72 hours, and the goal is that patients are getting to hospice 72 hours before they die. So is this reassessed or is this just a one-time assessment?

MS. SPENCE: The goal behind this was to get hospices right out of the gate working on pain for people that were uncomfortable and wanted pain management. It is similar to what Laura was saying yesterday. This is only at assessment. It has no implications for pain management down the road.

DR. LUNNEY: Sean?

DR. MORRISON: Carol, a question and just a comment. The question is: You presented data, the 19 percent or 15 percent, depending on the population, about groups. But I just wanted to follow up on the question about the quality improvement.

Do you have an understanding from your dataset about whether there's individual differences within the programs? I understand that the mean stayed the same, but the question is do individual programs change, and maybe I just missed it.

MS. SPENCE: Individual changes over --

DR. MORRISON: It goes back to the issue about the quality improvement aspect of this. One could say that the overall mean stays the same, but you are bringing in more and more hospices.

MS. SPENCE: Right, yes.

DR. MORRISON: So that there may be, but do you have data on whether you look at individual programs, whether that number changes over time for those individual programs?

MS. SPENCE: Yes, that is an excellent -- We have not specifically done that. I only know anecdotally. In talking to some hospices, some are like, yes, we do the measure. Do they do anything with it? Not much.

On the other hand, we have a couple of really good case studies where -- and I think I give the example in my material, where the implementation of this measure completely revamped their entire pain management program down to developing brand new competencies. I mean, it really got them investigating how that hospice was doing pain management.

There is your implications for moving forward. While the measure doesn't deal with it, the consequences can be such. So they put together an incredible communication process where there is actually email alerts done every time a patient is admitted where they answer yes to that question, and they go to the medical director who is going to be responsible for that patient, and the whole team then gets alerted that, yeah, this is something that we are now putting in.

As I said, and then in looking at how actually the pain management pieces, how well they are doing, it led to total revamp of their education around pain. So there's lots of different things that can come out of it. Right now, we haven't looked at the actual scores across individual hospices.

DR. MORRISON: I guess my comment is just a comment that follows from Naomi's question. You know, my wonky pain researcher hat says that, any patient self-reported measure is going to be open for bias in the clinical setting, and I just think we have to live with that.

MS. SPENCE: I think, if you are going to go back to the original -- pain is what the patient says it is -- then we are going to open ourselves up to some bias when you start measuring in that respect.

DR. LUNNEY: Eduardo, I wasn't able to detect you. Now I see how you hide.

DR. BRUERA: I think the crucial aspect of this is what you defined as the fact that we don't have the benchmarks, because as you very well point out, the patient has the right to call pain whatever they want to call pain, but that doesn't mean that that is not receptive input into the somatosensory pathway, and therefore, not treatable with pumps or opioids or sedatives, because what you can generate is suffering being treated with painkillers and adjuvants. And that would not necessarily be a quality improvement measure.

So part of the challenge is pain by nature is multi-dimensional, and when you summarize it to that nine, that number might be the lack of dignity touching off or might be your addictive disease that is causing you to call it that way.

So it needs a bit of a more thorough assessment. Not benchmarking a number is a very smart way to do this. Regrettably, if we have had this conversation, we might have results that are different today from yesterday with regard to the value of really determining some quality improvement measures without necessarily benchmarking that 30 percent or 20 percent or 60 percent is the right number.

So while the effort is good, in most cases coincide somatosensory nociceptive afferents with your complaint, there is a wide number in which that doesn't happen, and the risk of being tied to a number and a percentage is going to be as bad as not doing it at all.

DR. LUNNEY: I see most tents. I can't see them all. Are there any other on this side? That makes me ask the question, are we ready to vote? Okay.

The first question is: Is there a performance gap demonstrated?

MR. COLCHAMIRO: Seventeen high, three moderate, and zero low, zero insufficient evidence.

DR. LUNNEY: Is it -- oops, we are skipping this one.

MS. BOSSLEY: I think you can, actually.

DR. LUNNEY: Now that we are better judges of outcome measures, is this an outcome measure?

MR. COLCHAMIRO: Nineteen yes, one no.

DR. LUNNEY: Is there a quantity of studies to support the evidence of this measure?

DR. FINE: Can I ask a question? If this is an outcome measure, why are we voting on the quantity and quality? I thought we didn't need to if it was an outcome measure.

MS. BOSSLEY: It is a good point. I think we've got into the habit.

DR. LUNNEY: Could we skip this?

MS. BOSSLEY: I think you can skip -- Yes, we can go ahead and skip these.

DR. LUNNEY: And skip the next two. Okay. Is the measure itself reliable?

MR. COLCHAMIRO: Twelve high, eight moderate, zero low, zero insufficient evidence.

DR. LUNNEY: Is this a valid measure? Try one more time.

MR. COLCHAMIRO: Thirteen high, seven moderate, zero low, zero insufficient evidence.

DR. LUNNEY: So if there were disparities, would this measure find them?

MR. COLCHAMIRO: Nine high, eight moderate, one low, two insufficient evidence.

DR. LUNNEY: Is this a useful measure for public reporting or QI?

MR. COLCHAMIRO: Eighteen high,

two moderate, zero low, zero insufficient evidence.

DR. LUNNEY: And is it feasible to use this measure? Try again, please.

MR. COLCHAMIRO: Fourteen high, six moderate, zero low, zero insufficient evidence.

DR. LUNNEY: Finally, do we endorse the measure?

MR. COLCHAMIRO: Twenty yes, zero no, zero abstentions.

DR. LUNNEY: Good job. Then I think we are ready for measure number 1625, which is a measure developed by RAND. Is there anyone on the telephone from RAND who would like to present this measure?

OPERATOR: We do have Neil Wenger on the line.

DR. WENGER: Hi. I am here.

DR. LUNNEY: Good. Neil, we are about to discuss the hospitalized patients who die an expected death with an ICD that has been deactivated. Would you like to present an overview of that for us?

DR. WENGER: sure. I will be brief with this one. This is a new measure, which is in contrast to the measures that we have presented before, which have been through four different expert panels.

This measure went only through the ASSIST expert panel. It did follow the same RAND-UCLA methodology of linking process with outcome based on what the literature and clinical acumen would dictate.

This measure focuses on whether a patient admitted to the hospital who died after three or more days in the hospital and who have an active ICD in place have consideration of or deactivation undertaken prior to death.

The measure is a chart abstraction measure. It is a process measure, and it has been implemented only one time, and that one time was at a single institution at UCLA, and

I can't actually give you reliability data on it, because all of our reliability charts that were used -- none of them had an ICD in place. So I can tell you the reliability is perfect for detection of not having an ICD, but I can't tell you the reliability of the numerator.

There were 12 cases identified in our chart abstraction. Only three of them had consideration of turning off the ICD, which would indicate, at least preliminarily, the need for improvement in this area. But actually, there are more data from other publications, one that included 100 patients entering hospice with an ICD where only 40 percent of the patients had had consideration of turning off the ICD, and since then there have actually been further publications demonstrating that this is an area of great need.

Concerning validity, I can offer you the expert panel process as well as the relationship of groups of measures to important outcomes, but I cannot offer you any validity for this specific measure as far as process linking to outcome per se.

I think that this measure represents sort of an emerging measure for an area where there is lots of importance. There are a number of specialty societies that have indicated that this would be good process of care, and there is no doubt that, based on the emerging literature, that this is an area where process has not caught up with what those specialty societies would indicate.

I think that is probably about as much as I can present.

DR. LUNNEY: And presenting it from the evaluation perspective is Russell.

DR. ACEVEDO: Well, first let me just define better what the numerator and denominator were referring to.

First, as far as expected death, expected death is defined as the physician documentation, at least three days prior to death, that the patient's illness was terminal or that the patient had a grave prognosis, was receiving comfort care, was receiving hospice care, had a life threatening disease, or was expected to die. So that is the --

So the numerator will be patients from the denominator who had their ICD deactivated prior to death, documentation why this was not done, and then the denominator would be those folks with an expected death with an ICD in place.

The evidence was just mentioned. There really are no certification to a randomized controlled trial for this. Currently, the concept is that the outcomes would be better if the ICD were deactivated prior to death, and this is recognized as good practice.

I did pull up some of the clinical practice guidelines, and as mentioned that more and more have come out since this has been presented.

Looking at our review, we had six individuals who reviewed this, but the scores for high impact and opportunities were moderate. Perhaps the reviewers felt that the quality and quantity of the data was low to insufficient.

Both usability and feasibility scored three out of six as moderate, and two out of six as low. Four of the six reviewers voted not to endorse this measure.

As far as the comments, it was felt by the reviewers that this may be important to the individual, but this is a relatively small population, and efforts should be directed toward larger patient populations.

Also, they felt that conversations were very difficult to put in an EMR, and that this will require chart extraction to get the data. Also, conversations on the deactivation may have occurred in settings outside of the hospital, i.e., the cardiologist's office or the private physician's office.

The last comment was that a cancer patient with a defibrillator may not die or may not have V-tach or V-fib toward the end of life.

DR. LUNNEY: Do we have any questions coming from the panel? Sean, you were up first.

DR. MORRISON: I am actually -- Russell, since I am the co-investigator on an RO1 with randomized controlled trial about turning these off, I disagree with you about the possibility of doing that.

I am actually going to speak to some of the body of evidence, because I have been one of the people at Sinai who has been doing -- mentoring Nate Goldstein who has done a huge amount of work on turning off ICDs. So I can speak a little bit to that, because we have preliminary data that I can present to the panel as part of that process.

There were preliminary data -- I think Neil discussed this with the UCLA group -- in the grant that we put and was funded by NHLBI. The vast majority of people with ICDs did not have a conversation about these. There was a substantial number of ICDs that were not turned off prior to death.

To raise the question again, and it is painful, but I know these numbers, there are about 100,000 people who have an ICD put in every single year in this country. It is now estimated that there are probably at any given time 4 million people who are eligible under Medicare eligibility for that, and the growth of these devices is increasing dramatically.

So that it is becoming -- It is becoming a much more common device, and to talk about my anecdote, one of my patients who I saw last week had an enhanced pacemaker placed, and I asked him what was an enhanced pacemaker. He said, they put in something called a defibrillator as well.

So it is becoming very common, and I think that we are going to be seeing a lot more people who are dying of other diseases who have this incidentally. Anybody who has witnessed one of these going off in the setting of somebody dying, it is pretty horrific.

DR. LUNNEY: David, I think you were next.

DR. CASARETT: I will be quick. So I guess this is really more of a feasibility question, because we have certainly had these instances in my hospice, and they are bad, and they are awful, and they are memorable in an awful way.

I am thinking about a hospital that then would collect these data, and I imagine those hospital QI folks gong through every chart to find an ICD patient, and then looking in that subset to figure out what happened.

So my question is, yeah, I know 400,000 who have these, many more who will, but any idea what proportion of people who die in the hospital have an ICD in place, because I think that that number is very small. It would be a lot of work to look through these charts to get a small number of admittedly bad outcomes, but a small number. Does anybody know?

DR. WENGER: I can tell you, based on our data, that it was a little bit over five percent in 2006-2007, and in 2010-2011 we just looked at the same question, and it is a little bit over 12 percent. So it appears to be rapidly increasing, and it is becoming not that tiny.

Perhaps I can address one or two of the other points that were just made. The question of feasibility of the abstraction: Actually, published last year in the Journal of Palliative Medicine was the reliability of identifying expected death, and it was actually very high among a large number of nurse abstracters with Kappas of about .7 or higher.

The third point is that we are now working with at least one EHR to put this directly into their system, so that the feasibility of this is going to become electronic very rapidly.

DR. LUNNEY: Thank you, Neil. Robert, you had a --

DR. FINE: Do we have any data on how many patients with these devices actually do have a painful shock at death? We hardly ever -- At least at our shop, all politics is local, and generally, our cardiologists are turning these off before they call palliative care or hospice. I am just curious.

I can't think of any of our patients in the seven years at least when we have had a formal palliative care service where this has been an issue, where they have actually -- They have either already been turned off -- That is generally the case.

So I am just curious how much work it takes to find these cases, how often there is actually somebody who is getting shocked as a result of these as they die.

DR. WENGER: Can I address that? There may be someone there on the panel that knows the answer to that question, and I don't. But I would like to identify the difference between patients who die in the hospital with and without a palliative care consult, as those that are receiving palliative care consultations are very, very different than the vast majority of patients who die in the hospital who haven't received palliative care attention.

DR. LUNNEY: Russell, I think you have been waiting the most patiently.

DR. ACEVEDO: Sean, you said a randomized controlled trial. What are you randomizing?

DR. MORRISON: Can we have that

discussion at lunch, but I am happy to tell you about it.

DR. ACEVEDO: Okay.

DR. LUNNEY: Naomi?

MS. KARP: I just have a question about the denominator. I guess this is for Neil or Sean or one of the medical experts here, and it is just about the definition of expected death. I just want to make sure that that is not too broad, because it seems to include a lot of different possibilities here without any time limitations.

I am not necessarily advocating for anything different, but I just want to make sure that we are not having an over-broad category of patients here. So does anyone have anything they could add on that?

DR. LUNNEY: We have heard about the reliability. Do we have any evidence of the validity of this measure of expected death?

DR. WHITE: Can I just a sort of a

-- My concern is the same. I think this may be a too broad definition of expected death. I think a lot of us have in our mind the picture of a patient who enters hospice who is dying, and it is very clear where this patient is going and, of course, that patient could have their ICD turned off. But this definition allows for the patient -- allows for and would count as an expected death the patient who comes to the inpatient setting with V-fib who has got stage 4 CHF, who comes with V-fib who gets appropriately shocked and who dies, who wanted to get cured and treated or at least treated for their V-fib. That would still be a ding.

So I worry about this. Yes, if the patient came in and there was no documented conversation and they have -- So I worry about the life threatening illness part of that definition of expected death. It seems to raise a lot of problems that we might be comparing apples and oranges within this very broad denominator.

DR. LUNNEY: Sean?

DR. MORRISON: Just answer to one of the questions that I asked before. Nate's work, which I couldn't remember the numbers which I pulled up, although the data are all going back to 2004, when he looked at them, they found that in 27 of 100 cases somebody had discussed deactivating an ICD. It hadn't been discussed in the others, and that 10 patients actively received a shock when they were dying from their ICD at the end of their life.

Doug, I think the measure here is a conversation, not whether it was deactivated. Am I not --

DR. WENGER: That is true.

DR. MORRISON: Is that true, Neil? But I think it is about the report of a discussion, whether somebody would want their ICD deactivated in the setting of a potentially expected death. Is that correct, Neil, or am I not reading the measure right?

DR. WENGER: Correct. No, that is right.

DR. LUNNEY: I heard Doug's challenge to be the validity of the denominator, not the numerator. Am I mishearing you, Doug?

DR. WHITE: You are right. The denominator question is the one that I first asked. Sean's question was separate. If it is really about conversation, I would ask why is it not just -- Why is the measure not hospitalized patients who have a conversation about ICD before death, not -- There is some normative judgment being made here that you just definitely shouldn't be dying with it.

What I am wondering is would this just be better stated as did you have the conversation about it.

DR. WENGER: That would better reflect the measure, that title, and the numerator definition is the same as the title.

DR. WHITE: The only reason I raised this life threatening illness thing as part of the denominator is that there are some times when you come in with CHF, and the right thing to do is to keep your ICD on, and you can have the conversation and die with ICD on, and that is okay.

I think that there might be a relatively large proportion in the inpatient setting for whom that would be the case.

DR. LUNNEY: Well, the numerator statement is an "or" statement. So it is a deactivated ICD or documentation of why it was not deactivated. David?

DR. CASARETT: I am having trouble getting my head around this, but I guess (a) I would urge us not to, on the fly, suggest changes in the numerator because, obviously, a lot of thought has gone into these, but also related to that suggestion a discussion rather than deactivation or a discussion, it seems to me that you would get dinged if the discussion happened in the cardiologist's office.

Patient comes in. It is deactivated, but there is no discussion. So you get dinged. I didn't quite think through that, but I think that is an argument for not making these sorts of hasty decisions now.

DR. WENGER: Right. Actually, the way that you described it is precisely how the measure is implemented. The reason that it doesn't say discussion and deactivation is that the way that medical record documentation often works is that, if an action is taken, there is no description of the discussion, but if the action is not undertaken, there often is a description about the discussion that led to the choice not to take the action.

This is true across the board for depression treatment on and on. So a deactivation itself would satisfy, but the vast majority of cases where the measure is satisfied are documentations about discussions not to deactivate.

DR. LUNNEY: We still don't have clarity. Doug?

DR. WHITE: So now I am just raising it. Does it seem to the group then that -- I am thinking out loud about this, just keeping in mind this inpatient class who comes in with the goal of getting -- who come in with V-fib and an ICD, for example. Do we all still think that that should just be a routine question, even when you are coming in, in the acute setting, for cure or at least resolution of the acute issue is the goal, that you still have the conversation?

I could yes to that, I think, but that is kind of what we are saying here, is that even when it is pretty clear that the patient is coming in with a quickly reversible thing, you still have to have the conversation.

DR. LUNNEY: I think we have to be careful here, folks. We haven't had a benchmark that absolutely everybody should have this conversation or this deactivation, and some of our measures have those cases where we know from practice that it is not feasible to do what we see as a quality measure trying to measure.

So I think we just have to be careful to try to -- Many of our side conversations are about the difficulty of being even-handed in our voting and even-handed in the standard that we are using for these measures, but we have had the conversation several times over the two days that some of these measures measure a complex thing, and that we know that 100 percent is not going to happen, and we don't know what the right percent is that should happen.

Sean, you had a question?

DR. MORRISON: A clarifying statement to Doug, because I admit, I do come from it as a bias, because I am getting NIH funding to actually do research in this area, and think about this, and certainly have published in this area.

I do want to clarify one thing that you said, Doug, because I think the population that you are specifically describing is not what is in the denominator population that is put forward to us. The denominator population is hospitalization of adult patients of at least three days duration that ended in expected death.

Expected death is defined as physician documentation at least three days before death that the patient's illness was terminal, whether the patient had a grave prognosis, was receiving comfort care, was receiving hospice care, had a life threatening disease or was expected to die -- was expected to die.

I would suggest to you that anybody who has a life threatening disease -- and I think the data, both from the advanced care planning literature and particularly the work in the focus groups that have been done with patients with ICDs, suggests that people, the overwhelming majority -- no, it is not 100 percent as any benchmark -- do want to have that discussion. They may not want to have it turned off, but they at least want to have that possibility raised.

So I don't have an issue with the denominator population based on that, based that the measure is documentation of a discussion, and I do think there is a strong body of evidence to suggest that our patients do want to have that discussion.

DR. LUNNEY: Are we at a point of clarity to vote? Oops, no, we are not. Kate?

MS. O'MALLEY: Just a feedback or guidance. Since the numbers of people impacted by this are so small, does that -- I mean, I know we will have to make our own decisions. It just seems from the discussion that some of the information isn't really there, and the evidence isn't as predominant as might have been in other measures that we have considered, although I certainly agree that this should be done.

I am just wondering, based on the conversation that has happened to date, do people feel that we are ready to vote on this measure, or would more information be helpful, given the additional studies that you are involved in, Sean? I just would like some guidance on that.

DR. MORRISON: You know, Kate, I think you are right. We are not going to have data for another four years that we are going to be able to submit. I can talk to you about the preliminary pilot data that we did as part of our RO1 submission. I can talk about what has been published, which Neil actually, I think, has cited in the work here.

I think your point is extremely well taken. What is the right percentage? We have seen a doubling in the prevalence of these devices over the past five years. Now one in 10 people who die in the hospital has one in place. It was five percent three to four years ago.

You are right. We all have to make an individual judgment of what is the right prevalence, but it is 10 percent right now. It is one in 10.

DR. LUNNEY: Naomi?

MS. KARP: At the risk of beating the dead horse even deader, I guess I just want to throw out one more time: Is there a risk to patients who have a life threatening disease but could potentially live for years and are coming into the hospital for some -- I don't even know hypothetically what it would be, but something unrelated to this device -- that by having a conversation, they might do something that would shorten their life where they -- Is that a risk or not? No? Okay, good. Thank you.

DR. BRUERA: I guess one of the questions is the benchmarking. That is, some things do happen that you don't want to happen, but in the previous discussion we addressed the issue of controlling pain or failing to control pain, and those events do happen.

I think the question is, in one case are we trying to figure out things that are going to be a zero percent or are we trying to figure out the frequency with which some events occur, because that would give peace of mind to a lot of us that some situations in which somebody will die with advanced disease and will have the defibrillator go several times might happen, and that might be quite okay. But when there is a consistent trend for things like that to happen, then you have the C-section scenario of the person and the team perhaps not performing at the best level.

So I am not sure that we are tuning ourselves in the same way for every single question, and in some cases we seem to tune very high and in other cases we seem to accept the fact that it is the benchmark that will be defined over time, not that we are defining it right now.

DR. WENGER: Is it appropriate for me to respond to that?

DR. LUNNEY: Neil, yes.

DR. WENGER: I don't think the question is whether it is okay for people to get defibrillated numerous times before death. I think that what this quality indicator is trying to get at is whether that is okay without a discussion having occurred.

DR. LUNNEY: I don't want to open my mouth and stir anymore mud up from the bottom of this water that might be settling out. Is there some sense that we can now vote? Good.

The first vote: Is there a performance gap?

MR. COLCHAMIRO: Ten high, 10 moderate, zero low, zero insufficient evidence.

DR. LUNNEY: So is this measure a health outcome?

MR. COLCHAMIRO: Three yes, 17 no.

DR. LUNNEY: All right. Is there a quantity of studies that support the need for this measure?

MR. COLCHAMIRO: Three high, six moderate, nine low, two insufficient evidence.

DR. LUNNEY: All right. There are not a lot of studies. Are those that support this measure of high quality?

MR. COLCHAMIRO: Five high, nine moderate, three low, three insufficient evidence.

DR. LUNNEY: All right. Of those two measures of questionable quality, what is the consistency? Please try again.

MR. COLCHAMIRO: Ten high, seven moderate, zero low, three insufficient evidence.

DR. LUNNEY: We passed the bar? yes, we did. We go on.

MS. TIGHE: Heidi, I don't know if it did.

MS. BOSSLEY: It did. So let's go through the slides again, and we will walk through it. I was tracking correctly.

So the quantity was, I would say, moderate to low. But then if you go to the quality, the majority said moderate. Right? Then if you go to the consistency -- so if I look at that, it would pass it, I think, because again it says, if it is low quantity, moderate to high quality, moderate consistency, then yes as long as it is judged that additional research is unlikely to change the conclusion that the benefits to the patient outweigh harms.

I think the conversation we had was that you all agreed to that. So, yes, it did.

DR. LUNNEY: Okay, now we are looking at the measure properties itself. Do we evidence that it is a reliable measure?

MR. COLCHAMIRO: Five high, nine moderator, three low, three insufficient evidence.

DR. LUNNEY: Do we have evidence that it is a valid measure?

MR. COLCHAMIRO: Five high, seven moderate, six low, two insufficient evidence.

DR. LUNNEY: If there were disparities in care, would this measure capture it?

MR. COLCHAMIRO: Seven high, four moderate, one low, eight insufficient evidence.

DR. LUNNEY: Is this a usable measure for either public reporting or quality improvement -- or?

MR. COLCHAMIRO: Eleven high, eight moderate, one low, zero insufficient evidence.

DR. LUNNEY: We have a question?

DR. NEE: Yes, just sort of a quick question. Could we go back to the previous result? It seemed to me that, based on the results of that, we still have a fair amount of confusion in the group.

DR. LUNNEY: For the scientific acceptability?

DR. NEE: The one where it was equally high as it was insufficient.

DR. LUNNEY: The disparities?

DR. NEE: Disparities.

MS. BOSSLEY: It is not something that needs to be met. It is one that, especially if it is a new measure, I wouldn't expect us to necessarily have this information, but at the time of maintenance in three years, I think we really would like to see them having tracked the disparities question.

DR. LUNNEY: Okay. so is it feasible to use this measure?

MR. COLCHAMIRO: Seven high, eight moderate, five low, zero insufficient evidence.

DR. LUNNEY: And overall?

MR. COLCHAMIRO: Thirteen yes, seven no, zero abstentions.

DR. LUNNEY: I believe we have reached the point where we may go get some lunch. Oh, public comments? Thank you. Are there any people in the audience who would like to make a comment on our session this morning? Anthony, do we have anyone on the phone that we would like to invite to make a comment?

OPERATOR: All lines are open for public discussion.

DR. LUNNEY: So this is a time when, if there is anyone on the phone who would like to make a comment about the session this morning, they would be most welcomed. They must have heard our stomachs growling. We have 15 minutes to get some lunch and come back, and we will talk and chew.

(Whereupon, the matter went off the record at 12:17 p.m. and resumed at 12:33 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

12:33 p.m.

DR. LUNNEY: This next measure is the family evaluation of hospice care. This is the moment that Joan has been waiting for.

DR. TENO: I am going to do it here, if that is okay with you.

I am a pure academic, and by definition a pure academic cannot go anywhere without PowerPoint slides. So I did submit PowerPoint slides. I was wondering if we could project those PowerPoint slides, please. I am more of an academic than a doctor at this point.

DR. MORRISON: She says that, but every time I call her she is seeing patients.

DR. TENO: I guess you just have to stop calling me on weekends, Sean.

So if you could just make it a little bigger. Okay. Can I have the next slide.

First of all, I realize that I

probably caused some confusion with the committee, and I apologize for that. In submitting both measures, we were trying to be responsive to the precedents that Sean led and put together, and our intent really was to have an earlier version of the FEHC available for the potential to be used in evaluation of accountable care organizations or bundling in payments.

So what I would like to do is just sort of walk through with you both measures to really highlight what the joint development was and what the differences are, and we are going to go out to take questions.

Both surveys are based on focus groups that review guidelines, validation studies with tests/retests, national survey with evidence of discriminate validity on last place of care.

The FEHC has been adapted to self-administration with the mode test and some elements of repeat validation, and with both we tried to create as much as we can parallel creation of a zero to 100 composite score. Next slide, please.

Unlike what you have talked about a lot this morning is this is patient and, specifically in this case, bereaved family members perceptions of the quality of care. So one of the things that we have to think about is how do we get evidence of what is important from a bereaved family member's standpoint?

In doing this project, we took an analysis of 36 focus groups that were part of a Robert Wood Johnson Foundation values project, and then we supplemented those 36 focus groups with six focus groups of our own that looked at various settings of care.

We currently have just completed 16 focus groups from five regions of the country, and the initial survey was developed with Director Jack Fowler of the UMass CAHPS team, with the goal that each question asked only about things that are important to the quality of care for which the bereaved family member, we feel, would be the best source of information. Could I have the next slide, please?

The conceptual model that we developed was that high quality care at the end of life is when health care institutions provide the desired level of symptom palliation and emotional support, treat the patient with respect, promote shared decision making, attend to needs of the caregiver for information and skills in providing care for the patient, provide emotional support to the family before and after the patient death, and coordinates care across settings of care.

This model was based on conceptual guidelines, focus groups, and an expert panel.

One of the things I gave Lindsey this morning, in addition to giving a copy of my slides for you to take home and use as liners for your bird cage, I have also given you the table where we cross-walked the conceptual model to the NQF preferred guidelines, and started thinking the process of that linkage between structure, process and outcomes, and then showed you some of the actual survey questions from either survey.

We have tried to be very careful, realizing that we started on this in 1999. The NQF preferred guidelines came out some later. So our goal is to continue to grow the core instrument and make sure that we are staying with NQF and the National Consensus Project. Could I have the next slide?

So we used family as expert witnesses and to report the care that is delivered to them. Only three questions of the FEHC asks the family to act as a proxy. I think we are all well aware that the World Health Organization defined hospice and palliative care as the unit of care of that patient in the family.

We have chosen to focus on the

family just based on our experiences. It is too burdensome to try to interview these people in the last week of life, and also it is difficult to predict when that last week of life is. So that you get people at different disease trajectories.

So let me just sort of review the reliability and validity testing. Now the reliability and validity testing is really applicable to both instruments. We developed the survey. We did a test/retest among 29 persons. We dropped some questions based on them not meeting a satisfactory kappa.

We did an initial validation study with 156 persons, i.e., bereaved family members from hospice, nursing homes, and hospital. We reviewed our work on validation among 1380 persons. That was part of a national study. In that study we published in JAMA, we demonstrate some discriminate validity by the last place of care.

In our work on the development of

this, we involved an expert panel. We have tried to test construct, criteria, factorial and discriminate validity. Next slide, please.

So with adapting to the FEHC, we simplified to skip patterns, and specifically made some changes to the spiritual support question. We now use the word hospice team. We have a different time frame. The care instrument refers to last two to seven days of life, while the FEHC asks about the time the patient was on hospice.

We expanded the self-efficacy questions, which ask about the person's reading of their competence in doing the task by one, and we did mode testing comparing self-administration versus telephone administration, and did some revalidation of the self-administered survey, mainly consisting of internal consistency and criterion validity. Next slide.

So this is the graphical picture

of the overall zero to 100 FEHC score, showing a fairly good distribution, and I can give you specifics of the mean mode, etcetera, if you would like me to. Next slide.

Just to sort of highlight the differences between the various composite scores, just a minor correction: The scoring of the FEHC is based on having at least 14 out of 17 items present. We allow up to three items to be missing. We use sort of a mode substitution.

The constructs that are in the composite score is again providing the desired symptom control, emotional support, which is three questions, attending to the caregivers for information, skills is four questions on self-efficacy, two questions on whether you got enough information, knowledge about various symptoms, information to the family about the patient's condition and what to expect while dying, two questions.

We have an emotional/spiritual

support which has three questions, and coordination of care is three questions.

The Krumbach alpha is about .797. We have monitored coordination with a single rating of excellent, very good, fair, poor, and as you can see, when we take a look at the score, excellent has a 90.3; 76.2, good; 60.8 fair; 43.5 in poor; a 30.5 on a 100 score.

If a respondent says hospice is too late, the mean score is 79.8; at the right time, 87.1. So the care zero to 100 is based on 14 items. It again has a similarity to the FEHC in that we have three items on providing desired comfort and emotional support, but we include a treat the dying with respect with one question, emotional support to the caregiver with three questions, information and skills to the caregiver with three questions about self-efficacy, three on information. Information continuity is one question.

Internal consistency measured by

Krumbach alpha is .80. Correlation was excellent, very good, fair, poor, .58. Discriminate validity here by last place of care: if you died in a nursing home, the zero to 100 score is 71.2, versus home with hospice services, 83.2.

You know, right now more than 1200 hospice programs are using it. We undertook a process in terms of our measurement maintenance where we did three rounds of focus groups with end users, specifically quality managers. We asked them to give us sort of the bad, the ugly, and the good.

We are very pleased in those focus groups. We heard some areas that we could -- they wanted to expand it. They wanted some changes to the reporting structure, but the FEHC has been adopted to be really sort of the cornerstone of operation, staffing, quality, and it is the foundation of what a lot of hospices are currently doing for their quality.

All alone, we have tried to publish our results, noting the variation. Just refer to a 2005 article of the FEHC from the voluntary submission where we show individual variation on item, composite scores, and I previously had shown you variation on the new zero to 100 score.

Then I am not going to go over this in detail, but we are very committed to updating this instrument. We currently have a grant. We want to expand the instrument to better cover the needs of the Hispanic population. We are going to have a self-administered Hispanic survey that will be developed. We have a cultural anthropologist working with us.

We are going to do some more validity testing. We are also going to go to the future of doing mode testing with community administrations.

I think one of our goals as this goes forward is to do some harmonization of both the care and the FEHC instrument. So, hopefully, I haven't confused you thoroughly by now. Really appreciate the chance to present both these measures on behalf of myself and, really, Carol who has been working with me now for eight or nine years as a partner with the National Hospice and Palliative Care Organization.

DR. LUNNEY: Thank you very much, Joan. David, are you not the person who is doing it from the evaluation perspective?

DR. CASARETT: I am, or I am not "not the person," which means I am the person. So, thanks, Joan, for doing all my work for me. That was wonderful.

I have a full hospice unit of patients who need to be seen tomorrow. So if you want to put on your clinician hat, you could see those patients for me, too. Much appreciated.

I won't go over the FEHC instruments, since you have already had the graduate level course in that. So let me just summarize the results and pull out three or four points that came out in some of the comments.

So nine raters. It was endorsed by seven. Eight reviewers gave it a high impact rating. Somewhat less enthusiasm, five thought it was important, and I think there were some questions there about whether this was actually an outcome measure, which, hopefully, we have laid to rest, but we can go back and visit that one more time, if we have to.

There was some uncertainty about the evidence to support the measure. About half, more or less, of reviewers gave it a high rating for quantity, quality, and consistency of evidence.

In terms of instrument properties, there was more enthusiasm about its reliability than its validity, than its ability to distinguish among disparities.

Seven, five, and two raters respectively gave it a high rating.

In general, reviewers thought it was usable and feasible, six and five, respectively, and as I said before, overall seven people endorsed it.

So given those differences in scores, I went back through and did a qualitative analysis of the comments and came up with four or five themes you might want to address.

A couple of people raised questions about nonresponse bias and the impact on scoring. A couple of people raised questions about whether this was measuring proxy versus -- proxy for the patient versus family experiences, which I think you just addressed.

As I said before, ability to discriminate between programs, given many of the scores being fairly high. Two people asked questions about, I think, about the use of composite ratings for public reporting, so not an issue of scientific validity but a question of what would get reported to the public eventually.

Then a couple of people had questions about the equivalent of case mix adjustment concerns, that there might be differences in the way that different populations answered these questions, and whether or not that should be included in adjustment.

DR. TENO: Okay. So I have five questions to answer in two and a half minutes. Right? No.

So why don't we just take them from five and go backward. So we have thought very hard about case mix adjustment. I guess one of the things I didn't present in my overview that I was amiss in saying is we have published several articles about the differences by each of the composite domains by race, and that was published. It was summarized much better in the Care instrument application.

It could have been -- We can beef up the FEHC part of the measure at maintenance to put in references and the actual findings of the article, but needless to say, we do see differences by race in terms of African American versus white, and also Hispanic versus white.

To give you sort of a magnitude, the differences on Hispanics, I think, is fairly close to three points on the mean between white versus Hispanic. African American, I think, is two or one. It is not as predominant.

I am sorry, David. What was the question about the composite score?

DR. CASARETT: It was not, as near as I could tell, concerns about validity that were raised. It was the question of what the public would like to see, which may actually be beyond the bounds of what we want to discuss here.

DR. TENO: I think that is a very good empirical question that needs to be addressed. From our standpoint, we provide people who -- hospices that voluntarily participate in the FEHC repository both sets of information. We provide them a composite zero to 100 score, as well as the individual items, as well as the individual domains. We try to provide them with benchmark data so that they can compare their performance and, hopefully, improve their performance.

I think proxy versus patient I addressed, and the issue of nonresponse bias. So in the past we have compared the Medicare claims files, hospice providers, with those people age 65 and older who participate in the FEHC.

We found that we underrepresent African Americans and have a tendency to represent larger hospice programs in the measure. So that is sort of the bad news.

The good news is, when you compare us to other voluntary programs, we have a participation rate of around 42 to 43 percent. If you look at what the HCAHPS is doing, it is like eight or nine percent. So it is always amazing that with one mailed out survey, we are getting a fairly good return on the survey.

Then you had a question about programs, David?

DR. CASARETT: A couple of the reviewers raised questions about the ability to discriminate among programs. Again, this is my synthesis of the qualitative comments, but I think it was getting into issues of high scores across the board.

MS. SPENCE: Well, we ran -- If you look at individual questions, there are some questions that hospices performed very well on as a group, and there are others where there is more variation, but for purposes of the submission, we focused on the composite score, the large composite, that zero to 100 that Joan was talking about, which incorporates 17 of the questions.

We found, actually -- We did announce a variance and found that there was a statistically significant difference. We had a minimum mean of 73.3 and a maximum of 96.3, and with a mean of 86.6, and again that was statistically significant.

So I think that is pretty good evidence that it can discriminate.

DR. LUNNEY: Are there questions from the panel members? Rick?

DR. GOLDSTEIN: So can I just present a scenario, because I would be very interested in hearing -- I am sure you thought about it. So I am interested in hearing how you see the FEHC affecting this issue.

Let's say you are in a medium size town, and there are two hospices. One says the FEHC is very important, and I am just going to go with my bread and butter hospice patients, and I am not going to take -- I am not going to try to get to higher risk groups.

Another hospice says, you know what, we are going to take care of everybody else. We are gong to take care of the black families. We are going to take care of the Hispanic families who -- and we are not any better than the other hospice. So our performance profile is probably going to be muddied by the fact that we are taking on those patients.

How does the FEHC -- Does the FEHC ding the people that sort of go at an at-risk population, and have you thought about that at all? I worry about this as a tool that is kind of a disincentive for more complicated patients, as well as not really addressing the disparities issue that is so problematic in hospice care anyway.

DR. TENO: So, you know, I think you are hitting on a really important question, which I am going to answer philosophically. The question is should you put race into the multivariate model?

I think the first thing, before I address the race issue, is we have looked at a number of factors that relate to the zero to 100 score, and we found age, who the respondent was, number of times they have contact with the person in the last week of life, really don't predict how they are going to respond.

The one thing that does really predict how someone is going to respond is whether they are black or whether they are Hispanic. So the question is, is there something inherent about someone's skin color or the pigment in their skin that results them in doing a different rating score or is there something about the actual quality of care that they are receiving?

I actually think there are differences on quality of care, and I wouldn't want to adjust away those differences. I would rather have those differences out there as targets for improvement and having hospice programs think about how do we reach this very important population, how do we take our services and provide it to them in a culturally sensitive manner?

So as someone who has thought about this, I have decided on not using a risk adjustment for this, and really trying to put it out there for the emphasis that we need to improve hospice care for this part. Carol?

MS. SPENCE: I am going to take the prerogative of doing double time on this, since we are presenting two surveys together. Just to build a little bit on what Joan said, I think there is something inherently different about risk adjustment, because you have complicated diagnoses, and if you are in a hospital and this is the patient population, very sick patient population you serve, versus if you survey predominantly black community, and you are getting lower scores on FEHC, I think it is on the hospice to figure out how to better serve -- how to meet better the needs of that community rather than going in initially and risk adjusting.

We don't know enough. We haven't looked into it. Right now with the hospices that are participating, you have to aggregate at a national level, and sometimes over more than just a quarter, to really have enough -- a big enough n in the minorities to really start looking at differences.

So I am not saying there is not work to be done there. There is, but as it stands right now, I don't think we have enough information to say that we are putting hospices at a disadvantage by not risk adjusting based on race.

DR. GOLDSTEIN: But isn't -- May I just follow up? But isn't there a converse to that, which is doesn't this end up being a disincentive for hospices to go out and pursue these more complicated patients?

DR. CASARETT: I can try to answer that. That was actually the point I wanted to make before. That actually really worried me when I was reviewing this, but the more I thought about it, t hat is something that, I think, applies to all the measures.

The pain measure that we just looked at, that could arguably create a disincentive to enroll patients with pain. So I think all of these measures, virtually all of them except if you have a very fixed denominator beyond providers' control, would be susceptible to that.

So it seems like at least my take on this is that is a concern with the FEHC. I am not sure it is a greater concern than it is for some of the other measures. That was the way I interpreted it.

DR. LUNNEY: I just wanted to make sure that the question about public reporting was answered. I kind of, I guess, in my own mind wonder about the ability for the public to understand the complexity of this measure.

DR. TENO: Sure. I have not been involved currently in efforts to public report this measure. It is being publicly reported in Florida right now. While I have given full blessing -- You know, I've put this measure out there to let anybody use this measure free of charge, as long as they don't sort of start selling it or doing something wholesale wrong with it.

So Florida is publicly reporting it. I know the American Hospice Foundation has worked with data to design a public report card which Shoshanna Sofaer has done some initial testing, not the composite but individual domains, on that. There, I believe, might be a paper being written.

When I corresponded with Shoshanna over the summer, she was not at the point where -- She told me some of the issues they encountered, but it really wasn't issues with the measure. It actually was the issue with the public understanding what hospice is rather than the actual measure itself.

I think, if this goes forward for public reporting, just like any of the HCAHPS, the Consumer Health Plan Assessment measures, there needs to be some empirical work done to report the measure in a such a way that it is understandable to consumers.

I can tell you, as someone who has really sort of thought about this over time, we really tried to capture information that we thought would be valuable to consumers, that experts would endorse, and we tried to really emphasizes the face validity of the instrument, so that when a clinician hears the measure, it has sort of a, well, we got to do better, has that sort of clinical face validity.

Can I sit here and tell you right now that I have done this research to design a report, have gone out and tested? No. I think Shoshanna is the right person with her extensive experience with the CAHPS team to test this, and I wholesalely support the efforts to go ahead and do that.

So in a way, I am not able to directly answer your question, but I am concerned about it, and we want to proceed forward.

DR. LUNNEY: No, you did. Thank you. Kate, you had a question?

MS. O'MALLEY: Just a comment. There is about 200 hospice providers in California, and about a third of them collect FEHC data, either through a vendor or through their own efforts.

The California HealthCare Foundation has a consumer facing website that displays information about quality of care providers, including hospice. So we are working now with Naomi's group at the American Hospice Foundation to test the feasibility of putting FEHC data on our consumer facing website using a lot of the groundwork that they have already done in their experience in Florida.

We are surveying hospices now to see if they would be willing to participate with us, knowing that this data would eventually be reported publicly. So the surveys in the field, and I think we have got, I don't know, almost 90 hospices. Almost half of hospices are responding to our inquiry of whether or not they would participate in public reporting.

So just to let you know, the marble is continuing to roll forward, and there is a lot of interest in this measure in the field.

DR. LUNNEY: Are there any other comments or questions from the community?

MS. NAIERMAN: I should clarify that, when we worked with Shoshanna, she did a whole lot of cognitive testing of a report card design that she developed for consumers especially, and she did come across the stumbling block that most people that she tested with, consumers, really didn't understand hospice and, if they did, they were convinced it is a place.

So we were convinced then that any public report that is issued or that is published should be accompanied or should include an educational component that she designed based on the misconception she had found, and she cognitively tested several times over.

So our report card is posted on our website, AmericanHospice.org, under the tab Report Card, and you will get an overview of the work that she has put into it, and you will have an interactive model of how it works.

DR. LUNNEY: Thank you. I don't detect any other -- I have a better set of eyes next to me. Oops, Michael.

DR. LEPORE: From the sense of disclosure of interest, I am appointed in the

Department where Joan is employed. I know this is being submitted by NHPCO, but I do work in the Center for Gerontology and Health Care Research where Joan also works.

DR. LUNNEY: Then seeing no more comments or questions, I think we are ready to vote.

MS. BOSSLEY: But just so everyone knows, you are only voting on the first measure, 0208, not both, because I think we need to -- unless you disagree -- probably discuss both of them, because one is maintenance. So it is a little bit different looking at the public reporting, etcetera. The other one is a new one for you.

MS. NAIERMAN: When you say maintenance, when the FEHC was originally endorsed, it wasn't in composite form. Is that true?

MS. BOSSLEY: I would have to go back and look. Karen, did you --

MS. NAIERMAN: Joan would know,

because she was --

MS. BOSSLEY: Joan will know. I don't remember.

MS. SPENCE: We hadn't done the composite at that point. So it was actually the survey, but since it was very clear that NQF isn't in the business of endorsing surveys, at this point we submitted the composite, because that was the way to have a single score that included the maximum number of questions.

MS. NAIERMAN: So just another clarification: When we are voting on a composite, is it by implication that we are voting for each of the questions within the survey?

MS. BOSSLEY: This is where it gets fun. Okay. So they have put forward it as a composite, not with the individual questions pulled out to be reported separately. So you are really voting on that composite score of this as a result of the survey.

So everything you look at should be framed within that. Is that answering your question or -- I'm almost. I feel like I am almost.

MS. NAIERMAN: Well, for example, for me it was a little difficult to figure out about the usability. From our experience, the usability is the individual component. Now will the hospice be available 24 hours a day, or will they respect my loved one.

A composite is not as usable. So the rating for me would be different.

MS. BOSSLEY: So I think, because that is what is before you, you need to evaluate the usability of the composite itself, but I think you also need to keep in mind you are hearing that there is work on how to make this -- I mean, you yourself said it -- to make this understandable to individuals, although it sounds like you are using the individual components for public reporting.

But again --

MS. NAIERMAN: Yes, we are. But if this FEHC measure, FEHC tool, has already been endorsed by NQF, including its individual measures, can we not take it for granted that it has been approved by NQF already? I am talking about the individual measures -- the individual questions.

MS. BOSSLEY: Right. So the measure is up for maintenance now. Now I think this was -- I don't remember how many years before when it was actually the survey that was submitted, but as was clarified, we don't endorse surveys. We endorse the measures that result from that.

So we are not endorsing the individual measures right now, other than how they construct the composite. So what you would want to have happen, and this is part of our composite framework that we typically show and we can distribute to all of you so you see it, is that the individual components within that composite must meet the criteria, must show that they meet all four things, which I think, again because it was reviewed previously and I don't think there has been any significant changes, I think we can say that that has occurred.

Those can be and should be -- you should be able to drill down and look at the results of those individuals, but before you today is the composite, looking at the rollup.

DR. CASARETT: Just a quick usability clarification or question. I understand that we are supposed to rate feasibility based on the composite, but my understanding is that hospices would get each one of those individual items. So each one of those items that make up the composite would be actionable.

It seems kind of unfair maybe to penalize this based on usability, if hospices are seeing everything. Does that make sense?

MS. BOSSLEY: I wouldn't penalize

this measure based on the usability at all, as long as you can drill down. It is clear that there is work underway to look at how, when you get to the point of public reporting, which is where we really want this to head -- how do you then display that information.

Again, we may not rate this high because of that, but I wouldn't in any way factor of that into your rating.

DR. LUNNEY: With that final clarification, we are ready to vote. Is there a performance gap that this measure addresses?

MR. COLCHAMIRO: Seventeen high, one moderate, one low, zero insufficient evidence.

DR. LUNNEY: Is this measure a process or outcome?

MR. COLCHAMIRO: Eleven yes, eight no.

DR. LUNNEY: All right. Now in terms of the importance to measure and report this, is there a quantity of studies to support it?

MR. COLCHAMIRO: Twelve high, six moderate, zero low, one insufficient evidence.

DR. LUNNEY: Does that body of evidence have quality?

MR. COLCHAMIRO: Thirteen high, six moderate, zero low, zero insufficient evidence.

DR. LUNNEY: And are there consistent results?

MR. COLCHAMIRO: Fourteen high, four moderate, one low, zero insufficient evidence.

DR. LUNNEY: In terms of reliability?

MR. COLCHAMIRO: Fifteen high, four moderate, zero low, zero insufficient evidence.

DR. LUNNEY: Is it a valid measure?

MR. COLCHAMIRO: Fifteen high, three moderate, one low, zero insufficient

evidence.

DR. LUNNEY: If there are disparities out there, would this measure capture it?

MR. COLCHAMIRO: Eleven high, six moderate, one low, one insufficient evidence.

DR. LUNNEY: Is it usable and understandable for public reporting?

MR. COLCHAMIRO: Ten high, nine moderate, zero low, zero insufficient evidence.

DR. LUNNEY: Is it feasible to use this document -- measure?

MR. COLCHAMIRO: Twelve high, six moderate, one low, zero insufficient evidence.

DR. LUNNEY: And finally, do we endorse this measure?

MR. COLCHAMIRO: Nineteen yes, zero no, zero abstentions.

DR. LUNNEY: All right. We could move from here to measure 1623, the Bereaved Family Survey. Do we have a developer to present this to us? Okay, we will skip over and move to 1632. Do we need to hear more from the developers? Okay, Naomi is the one presenting this.

MS. KARP: Okay. So you have heard Joan's detailed pitch, and so I won't go over any of the characteristics of it. I will just talk about the people who reviewed it.

Seven people reviewed the measure. The bottom line was that five of the reviewers recommended approval, and two recommended against. Most of the ratings in most of the categories were in the high and moderate range. So I am just going to flat the concerns and the questions that were raised.

It seemed to me that the main concerns of the two naysayers appeared to have to do with feasibility. One mentioned that it was not easy to implement massively. Even some who recommended approval noted that it was ambitious, that more is needed to understand feasibility and optimal implementation strategy.

Some of the other questions raised were: Someone said the zero to 100 scale may be too much information for the public. I am not sure I really understand that. So whoever said that maybe could explain it.

There was a question about whether the survey instrument might be too long. There was a question about what is the interval between the time of death of the patient and the administration of the survey, and is that too long of an interval? I am not sure whether that time interval was specified in there. I believe in the published study that you attended, you talked about a nine month interval, which seems like a long time.

Then someone noted the fact that, unlike the FEHC, which is just for hospices, this is meant to be administered in several different settings of care, and there was a question raised about whether hospitals would really be willing to devote the time needed to this, that hospitals were very different from hospices, and also that families of people who die in the hospital may not be as focused on the quality of the care at death as people with patients in hospices.

DR. LUNNEY: Would the developers like to respond to those?

DR. TENO: You know, actually, I am going to go last and get confused as usual. Our experience along with this national fielding of the survey we actually did 100 in depth narratives, and that is just not something we saw.

Irregardless of the site of care, the dying of a loved one is a sentinel event that really does impact. When it goes wrong, it really leaves an impact on people that you can recall those memories up to two years later.

We chose nine months for the study based on when we could get the death certificates. It is a research goal right now for us as part of the development of the FEHC to look at early versus late administration, look at stability and responses over time. It is on our agenda. We, hopefully, will have that answer to you by the time of the next measurement maintenance, and make more solid recommendations.

The feasibility: We agree with you that our intent and the reason that I have put this forward was at the request of the NCPRC. So that if we do have a demonstration going forward regarding accountable care organizations or bundling of payments, we wanted to have a measure out there that could capture the consumer's perspective; because you realize that all of these changes in how we finance our health care system, the people at most risk is the dying, because those are the high cost patients.

So our intent with this measure was to fill a void, because initially this was not something I was going to do. I was sort of drafted in doing it. I want to make it available as an ACO.

This is a measure that, if it gets picked up and gets used, I am going to try to harmonize it back to the work of the FEHC.

DR. LUNNEY: Are there any other -- Yes, Eduardo.

DR. BRUERA: Just for clarification, Joan, is this a copyrighted tool or is this going to freely available to everybody who wants to use it for their clinical programs?

DR. TENO: My life is online. Seriously, I have been very fortunate to work as a health services researcher for 25 years, and I really believe the onus on me is to give back to the field. All my work is done through either taxpayer money or public philanthropy.

So from the very beginning, I have put everything out there on websites, made the information accessible. The instrument currently is being used in Australia. I get these wonderful emails from the person in Australia who is using it.

I have let people sort of adapt it. It has been translated into various languages. I have done freely. I just think it is a service that we do, because all this work comes from really either federal or, in this case, philanthropic money.

DR. LUNNEY: Robert?

DR. FINE: Joan, a question. Is this just for veterans or is this for anybody, because it refers to veterans?

DR. TENO: No, that is 23.

DR. FINE: Ah, that is why I am confused. Thank you.

DR. LUNNEY: We are on measure 0208, the care measure.

MS. BOSSLEY: This is 1632, and 23 is the veterans, I believe.

DR. LUNNEY: Naomi.

MS. NAIERMAN: I have a point of

clarification. Picking up on what I asked earlier about the individual components, being that this -- I hope I am getting this right. This is similar to the FEHC. One is the mother of the other, and we are asked here to consider all the individual components and not the composite. Right?

DR. TENO: This is something I would like NQF guidance on, because obviously, the team struggled on how to write this up, and struggled with this whole notion of maintenance.

I did this in Seattle, and I decided to give you the maximum information, not only to give you the zero to 100, but to show you each of the subdomains that we have; because I actually think they are both important, and that as you go forward in public reporting.

If the NQF agrees, I would like to update, with Carol's permission without her hitting me -- I would like to update the FEHC to have a very similar -- you know, talking about how the FEHC really, in addition to having the zero to 100, also does have components, which I would like to make that information available, but she is not going --

MS. SPENCE: No, I mean, that is -- We had a telephone call, and we discussed that point -- we, NQF and NHPCO had a telephone call, and we were told very clearly that, if we wanted to do individual questions, there had to be a separate submission for each individual question.

So from a practical standpoint, that was also part of my decision to go with the composite. I don't know if there is a difference between that being a maintenance measure versus this being --

DR. LUNNEY: So this is something that I will say even as NQF staff, we have spent quite a bit of time and went through our Consensus Standards Approval Committee, to talk it through as well, because it is again an evolving art, I think, in measurement, looking at surveys.

So there are several ways that this could be put forward, and I think we can talk offline if there are ways that you want to further pull out either the domains or the questions that you feel that are relevant for public reporting and accountability.

What we have before us is the composite, which is the rollup, for both the FEHC and the care. Correct? I want to make sure I am understanding, because it is not necessarily clear.

DR. TENO: Sure. To be clear, Carol, listened to you -- I didn't. Carol gave you the composite. I actually gave you the composite and the psychometric properties of all the domains as well, and I provided you with all the validity testing that we have done with each of the items. But that is because I have obsessed about this for about 12 years now.

MS. BOSSLEY: Right. So we want all the information underlying the composite that shows how everything pulled together into that composite. So the individual, the domains, all of that, we would want. So what you have done is correct. I think both of you have done that.

What I think is the question that we may need to take offline and then bring back to the committee is whether just the composite that we have in front of us, because I think that is how the committee has been voting, and that is what we have looked at for both of them, then needs to be further broken out into either domains or something.

I guess part of what would be helpful for you all to think about is are there specific domains or questions within the survey that makes sense to pull out and be reported on their own.

That would be the part that, I think, you would need to tell us. It could be all of them, and then we will figure out with you how to help get that information into what we need.

DR. TENO: Personally, I will be very glad to work with you on that.

DR. LUNNEY: Are we clear on this?

MS. BOSSLEY: So you all know what you have in front of you today. In front of you today, I would say, is the composite. So it is the rollup of everything. You have discussed it for the first one, 0208.

Now you are discussing it for 1613. What we need to figure out, and we will do it with Joan and Carol, is whether or not we are going to bring back to you the individual questions or some domains that would be pulled out and actually be endorsed as separate measures that could be reported out on patient experience -- or family experience.

DR. LUNNEY: I just want to clarify that I am understanding the denominator correctly, and this is intended as a follow-back survey that could capture next of kin of anyone who died in any setting that was not traumatic.

DR. TENO: Non-traumatic deaths, and the next of kin has to say they were the person most involved in the care and would have, or did, make medical decisions.

DR. LUNNEY: Okay. Let us proceed with our voting. Is there something? Come on, Doug, hurry up. All right.

Has there been a performance gap identified to support the importance of this tool?

MR. COLCHAMIRO: Fourteen high, five moderate, zero low, zero insufficient evidence.

DR. LUNNEY: Is this measure and outcome measure or a process measure? Is it an outcome measure?

MR. COLCHAMIRO: Eight yes, 11 no.

DR. LUNNEY: All right. Is the

evidence to support the importance of this measure -- is there a quantity of studies available?

MR. COLCHAMIRO: Nine high, nine moderate, one low, zero insufficient evidence.

DR. LUNNEY: Again, in the evidence that supports the importance of this measure, is there quality evidence?

MR. COLCHAMIRO: Eight high, 10 moderate, zero low, zero insufficient evidence.

DR. LUNNEY: Among the studies supporting the importance of this measure, is there consistency?

MR. COLCHAMIRO: Ten high, nine moderate, zero low, zero insufficient evidence.

DR. LUNNEY: Now looking at the measurement properties, is there evidence that this is reliable measure?

MR. COLCHAMIRO: Eleven high, eight moderate, zero low, zero insufficient

evidence.

DR. LUNNEY: Is there evidence supporting the validity of the measure?

MR. COLCHAMIRO: Nine high, 10 moderate, zero low, zero insufficient evidence.

DR. LUNNEY: If there are disparities out there, will this measure catch them?

MR. COLCHAMIRO: Ten high, nine moderate, zero low, zero insufficient evidence.

DR. LUNNEY: Is this a useful measure for public disclosure or quality improvement?

MR. COLCHAMIRO: Nine high, nine moderate, zero low, one insufficient evidence.

DR. LUNNEY: And is it feasible to use this measure?

MR. COLCHAMIRO: Seven high, 10 moderate, two low, zero insufficient evidence.

DR. LUNNEY: Finally, do we

endorse this measure?

MR. COLCHAMIRO: Nineteen yes, zero no, zero abstentions.

DR. LUNNEY: I think we are ready for our last instrument that we are going to deal with today, number 1623 from the PROMISE Center, the Bereaved Family Survey. Do we have a developer? Is the developer available?

MS. TIGHE: Anthony, is there a developer from the PROMISE Center on the line? I think the name is Heim Lu.

OPERATOR: We do not have that company on.

MS. TIGHE: Thanks.

DR. LUNNEY: Then, Rick, over to you.

DR. GOLDSTEIN: So the BFS is a comprehensive measure intended for use by the VA for quality improvement in the care of veterans who die in inpatient VA facilities, and it assesses families' perceptions of the quality of care for those veterans who -- and the care was during their last month of life, with a follow-back study administered to identified next of kin six to 10 weeks following death, and its elements focus on communication, emotional and spiritual support, pain management, and personal care needs.

It also has sections that ask about the presence of PTSD at end of life, which is of interest to the VA, and also family awareness of entitled benefits.

Its numerator is completed surveys, completed being 12 of 17 items, with optimal responses to a global assessment of care score, with additional breakdown of tested elements. The denominator is the total completed surveys from the family member with the VA inpatient death.

They excluded patients without identified or contactable next of kin and acute fatalities, and six of us completed evaluations.

In summary, it provides a performance gap. We all rated it highly. We were split about whether this was an outcome measure, and if it was, it would have scared us from the evidence presentation. However, the measure designer didn't feel that the structure was -- that structure process/outcome relationship applied in this case, but then later cited the measure's discriminate validity and showed a relationship, a testable relationship.

There is only one study cited, but the measure is based on FATE and FATE-S for which there was more evidence, and you can see from the scores that we were all over the place in assessing quantity, quality and consistency.

It has been tested, pilot tested. The measure has been pilot tested with suitable operational characteristics. There is general agreement that the measure was usable, but less agreement that it is feasible, mostly due to the fact that it is an add-on process.

In terms of specific comments, it was noted that crucial aspects at the end of life care are not included in the measure, things like advance care planning, coordination of care, and family burden.

The designers have noted that the care of veterans without next of kin and families is more costly and tends to have poorer outcomes, and this measure does nothing to address that.

There was some question about the measure's use in the general population, and that concern led to the only exception to all other reviewers vis a vis it is a suitable measure for endorsement.

I think this is really just intended for the VA, and that they are asking for endorsement.

DR. LUNNEY: Given my side question, does the NQF address measures intended for specific populations, and the answer is yes.

MS. BOSSLEY: Right. Let me just add, though, I guess the other question would be is there anything that this survey perhaps addresses that couldn't be addressed by the other ones you have looked at in that population? I think, to have you have discuss that -- So some of the other ones kind of could be potentially applied to this population, perhaps not because of the group we are talking about. That may be helpful, to give us a little feedback on that.

DR. LUNNEY: Is anyone familiar with this to know whether there are aspects of the other instruments that were deliberately left out or anything added that was not covered by the other -- the FEHC or the CARE?

DR. GOLDSTEIN: There are some specific VA measures. So there are two questions about PTSD, and there are about three questions, if I remember correctly, about whether there were death benefits, burial benefits, and family benefits that follow.

I would also say that the language of this, which I think is the strength of the way the questionnaire is constructed -- It is a telephone survey, but it is very accessible, and I am just guessing that, if we tried to estimate sort of an educational level, it is at a lower level than what I reviewed from the others.

Then in terms of what wasn't there, things like coordination -- the same things I just talked about. Coordination of care and advance care planning, and family burden are not included.

DR. LUTZ: And you may have just answered this, and I maybe didn't understand. But you said that NQF can look at something that is just for a specific subset of health care, but then we also have the charge that, if there are two competing things, we are supposed to pick the better.

Does that mean this can be separate altogether or does it mean that one has to be the better?

MS. BOSSLEY: I would say, given the population we are talking about, I think that it is a unique enough population that it makes sense to have a separate survey and a separate measure.

So if you all agree with that, that is how the recommendation would go forward, that it is appropriate to have two.

DR. LUNNEY: Let's see. Tina?

MS. PICCHI: One concern I had was the small group of people that it was addressing. Only 27,000 of the veterans in 2000 who were in veteran facilities had this survey administered, and the other 77,000 who died were not included in this survey.

I didn't know if there was a -- I know that they have been trying to tie this as a quality indicator to the palliative care programs that they have implemented in the VA, and so it is specific to the facilities. Nonetheless, there is a large population of veterans' families who would not be surveyed with this.

MS. NAIERMAN: I think I could answer that partially. To the extent that VA facilities refer out to hospice, they may be -- veterans may have been captured in a hospice setting versus dying in the inpatient facility. So the 27,000 is a little bit deceiving. Of course, the rest of it may be that they didn't have a family member that can be contacted.

DR. LUNNEY: David?

DR. CASARETT: I think I can add to that a bit. although I am not familiar with all the details. Two issues, problems, I could see. One is I think it is very, very difficult to identify people who die outside of facilities. Some veteran who dies somewhere in the community, figuring out that they have died is difficult.

The other issue is I believe that NHPCO is developing a version of the FEHC or a module of the FEHC that uses some of these questions as a way of getting at some of those patients.

DR. LUNNEY: I am not sure I quite understood that, David. Are you saying, therefore, that in the future the FEHC or the CARE will make this not needed?

DR. CASARETT: No, I don't think so. I just looked at Carol who, I guess, is not allowed to speak. So I am going to channel Carol for a second, if I can.

So the FEHC is administered to people who die in hospice, some of whom are veterans. So I believe NHPCO is working to identify those veterans and ensure that those veterans get an additional module. That would be different than those veterans who die in a VA facility, a nursing home, for which I think the BFS is directed.

DR. LUNNEY: All right. You can nod.

DR. CASARETT: Carol can nod.

DR. LUNNEY: And presumably, the concerns in the VA facility are slightly different from the concerns that might be had if the veteran was out in the general population. So, okay. I think I am clear.

DR. GOLDSTEIN: I just wanted to make one other comment in favor of not harmonizing this too early. It has to do with the VA system. They don't have problems with access. SES is lower. Increasingly, the demographic mix is traditionally a much harder group to capture in our studies.

So I would be very interested to see what they find, and I think that they should be able to pull out a lot of interesting information.

DR. LUNNEY: Are there any other questions before we vote?

MS. O'MALLEY: I just had a

question. This is just more of a wonderment kind of a question of the purpose of seeking NQF endorsement for this particular measure, because the thought that comes to mind is that the VA is the federal government, and if you are thinking about NQF's endorsement as a launching pad for promoting uptake and promoting the quality and people's interest and engagement in quality processes and public reporting, I am just wondering why this measure would need NQF endorsement.

If it is the federal government, you just say do it, as your measure.

MS. BOSSLEY: Well, I would have to ask the VA why they submitted it, but part of the -- The government is actually required through AHRQ and other legislation to use NQF endorsed measures wherever possible. I think that would indeed apply to the VA as well.

So my assumption would be that may be part of why you see this measure before you now. My other assumption would be a lot of people really look to that multi-stakeholder input that the NQF process has, and does find value in it. But again, I can't speak for the developer themselves, but I think that may be two factors why you do see the measure before you.

DR. CASARETT: I don't think I could speak for the developer either, but there is a rhetorical process that goes on in both convincing people that a measure should be used within the VA, given that right now there are something like 200 quality measures the VA use.

You need to convince people that you really need a 201st, and this costs money. As people have mentioned, there is a commitment by the VA to do this survey for every single inpatient death across the country, which is expensive, and it becomes easier, I think, to argue for that expense for an NQF measure. So I think it is those things, too.

MS. NAIERMAN: I just have a quick comment. Someone said earlier that this may not be totally feasible, because it is an add-on, but to Kate's comment, if they are told to do it, they are going to; and if the funds are satisfied, they will do it. So add-on or not, seems to me, it will be done.

DR. LIAO: So, David, can I just ask a clarification question. So the VA has 200 quality measures. They don't -- Each VA hospital does not have to implement or use all 200, right? They select which ones they are going to use?

DR. CASARETT: I am no longer in the VA, a factor for which I thank my lucky stars every time I wake up in the morning. So I don't know what the latest count was. I do know that, when I used to work in the VA, the question always was we are being held accountable for way too many things for us to keep track of; don't give us another one.

DR. LIAO: At least I know my

local VA -- the director gets to pick which ones sort of that they are going to use and report, for whatever their incentive program is. So, yes, they don't have to use -- I guess that is why --

DR. LUNNEY: However, I don't think it is our purpose to worry about what the VA is going to do with this. I think our purpose is to determine whether this is a measure that we think warrants endorsement by this panel as a measure that has substantial evidence supporting the need for something to measure and document performance gaps, has the liability and validity that we would expect of a measure that was to do that, is feasible and is useful. Is that fair?

How they choose to work with our endorsement or not is not for us to deal with.

So having preached, are we ready to vote? Is there a performance gap?

MR. COLCHAMIRO: Fifteen high, four moderate, zero low, zero insufficient

evidence.

DR. LUNNEY: Is this measure process or outcome?

MR. COLCHAMIRO: Ten yes, nine no.

DR. LUNNEY: All right. In terms of the importance to measure and report, is there a quantity of studies?

MR. COLCHAMIRO: Eight high, 10 moderate, one low, zero insufficient evidence.

DR. LUNNEY: Is there quality in those studies?

MR. COLCHAMIRO: Six high, 12 moderate, one low, zero insufficient evidence.

DR. LUNNEY: And is there consistency among those studies? If you have made up your mind, try again.

MR. COLCHAMIRO: Seven high, 11 moderate, one low, zero insufficient evidence.

DR. LUNNEY: So is there evidence that this is a reliable measure?

MR. COLCHAMIRO: Seven high, 10 moderate, two low, zero insufficient evidence.

DR. LUNNEY: Is there evidence that it is a valid measure?

MR. COLCHAMIRO: Seven high, 11 moderate, one low, zero insufficient evidence.

DR. LUNNEY: If there were disparities, would this catch it?

MR. COLCHAMIRO: Eight high, nine moderate, zero low, two insufficient evidence.

DR. LUNNEY: Is this useful for public reporting or quality improvement?

MR. COLCHAMIRO: Twelve high, six moderate, zero low, one insufficient evidence.

DR. LUNNEY: Is it feasible to use this?

MR. COLCHAMIRO: Eight high, 11 moderate, zero low, zero insufficient evidence.

DR. LUNNEY: And overall do we endorse it?

MR. COLCHAMIRO: Nineteen yes, no zero, no abstentions.

DR. LUNNEY: I believe that allows

us to set these things down for the day. Before public comment, we are going to have the framework discussion.

So at this point, I will turn it over to Heidi for a discussion of the report's framework.

MS. BOSSLEY: So just one thing we wanted to -- and this will be very brief, because again I think we will continue to have conversations about this. Did send around the preferred practices, that report from the last maintenance -- or endorsement? I don't think we did.

So what we will do is we are going to send around the last projects that we did related to palliative and end of life care. In there, it looks at both measures as well as defining preferred practices, and created some domains with which you would kind of frame the whole spectrum of care in this arena.

We would like to see if you wanted to still think that is relevant. Do those domains continue to capture what we are looking for, and then we will begin to structure the report around that. If not, though, we wanted to get your feedback on that, because there are several groups out there currently trying to also kind of capture how care in the domains with which care is provided in this area.

So, for example, CMS has a technical expert panel currently looking at developing measures, identifying gaps, and they have come up with several domains, as well as the Long Term Quality Alliance has also done the same.

So we will distribute all that information to you, and then would like to propose that we use the same structure and format that we did in the last report. We want to make sure that we are still relevant to how care is delivered now.

So just a heads up that that will be coming to you on how we re thinking of framing this report.

The other piece, though -- and I am going to let June run this -- is we always like to spend some time after you have gotten into the weeds on the measures to kind of step back and take a look and see, within care for palliative/end of life, where are we in the set of measures that you would like to see to be able to show the quality that is being delivered to patients, and what gaps are there? I suspect there are a lot.

So I would like to just begin a conversation today, because a part of what we include in all our reports is here are the measures you put forward, but here is where we still need measurement and quality improvement focus.

DR. LUNNEY: I actually would like to just say it is easy to get caught up in the sense that we are packing up and, yes, okay, we are through with our laptops maybe, but please don't be through with your brain, because I think this section of our meeting could be very valuable in terms of our stopping now.

Go up to the 10,000 level, look over the set of measures that we have evaluated in the past two days and those that we have endorsed. What is missing?

All right. Doug was the quickest that I saw, but maybe I was just looking straight ahead.

DR. WHITE: I will just start with one to get the ball rolling. Last night I looked at some of the consensus guidelines specific to the acute inpatient setting related to decision making about life support and goals of care.

One of the things that struck me is that, especially in the critically ill population and the acutely ill population, there is a need to assess not just whether there is this one conversation, because we know in the inpatient setting often it is the act of talking early and talking often that is important.

In fact, a lot of the good work that Judy Nelson has recently been doing is about how can we get people to talk serially over the arc of an illness in the inpatient setting, and she has proposed that, and recently they tried to use that as a quality measure in a project, I guess, in Rhode Island was one of the big things, and it just failed miserably. They weren't able to get people to talk more serially over the course of a hospitalization with the patient or the surrogate.

So that would be one, thinking about ways -- Can we document? Can we have a measure that talks about serial conversations, maybe within 48 hours and then within weekly within the ICU?

DR. LUTZ: I just wanted to echo what Eduardo said earlier. I think part of the concern I have as an oncologist coming here is that there are a lot of issues, and I sort of feel like we are putting the computers back in the bag, and because there was sort of an absence of someone to carry the ball for the ASCO issues.

You know, we've got to the point where we said we had one or two favorites, but that doesn't give me a lot of closure. It leaves me believing that, unless someone can bring those to the finish line, we have not taken the opportunity to help ourselves through what is a very difficult time in oncology, as oncology struggles a great deal with what to do with end of life care.

There seems to be a wide schism even in the societies. We can't predict that well when someone is going to die of cancer, but we can more so than many of the other things that lead people to hospice. We still don't have very good measures.

So I feel a little empty that those didn't get done.

MS. BOSSLEY: Just to say that you are going to have more conversations on this. So please go away feeling better. We are not done yet.

DR. FINE: So my hope is that at some point tools might be developed that look at the quality of care at the end of life for all patients and not just those seen by palliative care and/or hospice teams.

I understand the need to focus on the quality of what we who are palliative care practitioners or hospice practitioners do, but it strikes me that we will never see all patients who face the end of life or there is certainly a period of time where we are not yet involved, and yet there are qualitative issues.

It seems to me the tools in general seem to place in their denominator in general patients that are kind of already identified as palliative care appropriate.

DR. LUNNEY: Thank you. You might

straighten your table cards, if it is appropriate, only if it is appropriate, so that I can keep some track. I am going to continue, if you don't mind, going around the room, and I will come back. So, Kate, I think you are the next whose tent I see.

MS. O'MALLEY: One thought that comes to mind as I look back on the work that we did, it seems that our measures sort of conform with the fragmentation that we have in our health care system.

In relationship to Bob's comment, would really encourage the development of measures that go across the continuum of care. I think specifically about the outpatient setting like medical groups in California that deal with elders or nursing homes.

You know, in California 22 percent of frail elders die in nursing homes, and when we look at discharge data from nursing homes, 38 percent of nursing home discharges are hospital admissions, and hospital deaths --

about 20 percent of them were preceded by a nursing home stay.

So there is an enormous amount of churn out there, and I think people are very well aware of that. I don't know yet exactly what the role of quality measures can play, but finding things to support in studies and to look at settings of care, particular nursing homes, I think, would be very valuable to have a higher degree of quality scrutiny and ability to measure in those settings.

DR. LUNNEY: Thank you. Eduardo.

DR. BRUERA: I think our challenge is going to be to find very hard outcomes, and basically, as you know, if I fly tomorrow and my plane lands, 80 percent of the people are going to be delighted with the flight, even though the food was awful, it landed three hours late. We are alive. We landed, that's fine.

Mom is sick. Mom got admitted. Mom died. So the question I have is it is very hard sometimes to pose the question on the people who are the recipients or who are suffering through the problem at that point, and there is nothing inherently wrong with that, but we need to go much more hard in the outcomes.

There are things that people are undergoing that should be the real outcomes. The reason why we are going bankrupt is not because we are having some chance -- it is because of things we are doing and the results of those.

So I think we need to step back a little bit and look at exactly what the things are that are low hanging fruit, that are very hard outcomes.

I would like to echo some of the issues, that I don't go home with the feeling that asking this question or asking that question is really going to change end of life care in this country. So measuring hard benchmarks that are out there and getting the consensus that we don't have to feel bad or paranoid about them -- they are benchmarks that will be defined, the same as the C-Section was defined. But if we don't put those on the table, asking this or that question -- I ask questions to death to my patients, but I am not sure the answer is that is what we do to people, that we need to put back on the table.

DR. LUNNEY: And I am going to take my place around the table, and particularly because -- oops, Naomi.

MS. KARP: I guess I want to focus on advance care planning and documentation. Admittedly, I haven't seen the NQF advance care planning measure. So I don't know what is in it.

I guess, to Doug's point, I want to -- and I think that you were talking more about the hospital setting, but I agree with you about advance care planning is not a one stop shop, and so we should look at a measure that goes across a longer time frame. But I also think we should do that in more settings of care. So we should expand it to outpatient medical practices, nursing homes, home care, etcetera.

I guess the specific one -- I haven't really thought about this before, but we were talking about POLST before, and I wonder whether the use of the POLST could be amenable to actually being a quality measure.

In many states, it needs some kind of legislation or regulation, but it is still -- even without that, it is a form and a protocol that can be used within a facility, and is not necessarily dependent upon that. So that might be something very concrete and tangible that we could look at.

DR. LUNNEY: And what I wanted to bring up -- and it is ironic, because Naomi reminded me of a mutual friend who, if she was here, would certainly bring it up as well, and that is the fact that the National Mortality

Follow-back Survey is dead on the vine for federal funding, but at least one of our measures today had a very broad denominator that would have caught many more people who died than those that we saw it coming.

It does become a real issue, because some of the questions that you want to ask, you then can't turn around and change care, because if you didn't see it coming, you didn't know you were caring for someone who was dying. But most of the questions that we ask are questions that would indicate whether health care was of high quality, whether the person was headed to a end or not.

So I would like to say that I think we could develop some good mortality follow-back survey questions and encourage their use, even outside of the federally supported nationally representative survey.

Around the table. Did I skip you? I'm sorry.

DR. GOLDSTEIN: You did. That is

okay.

DR. LUNNEY: I don't do vertical versus horizontal.

DR. GOLDSTEIN: I think POLST is really something that is important, not just in hospitals but in terms of transitions in care, and also continuance of care, and it seems like it a ripe fruit that is ready to be plucked. So I would say that.

There were really no process measures related to communication of critically ill patients. So something like ICU family meetings might be an interesting benchmark to explore.

Then just from where I sit, there really are no measures either specific or attempting to address children and young adults and issues like minors with decision making capacity, the presence or availability of hospices with expertise to take care of children, or even the availability of functional services. OT, PT, child life educational support services in the community for critically ill children and families are things that I just feel are missing.

DR. SCHROEPFER: I have two areas I wanted to raise. The first is in terms of the role of culture. My research focuses working with medically underserved communities. I work with a number of tribes in Wisconsin, Amish, Hmong, so different communities.

In working with the communities around end of life, I have learned that how we measure things, we are not really going to capture what we are seeking with regard to some populations.

One of the examples I can think of is working with the Shakopee Nation in Minnesota, what came out of the research was that stoicism is so key for elders. So an elder can be in pain, and the chart that they use there, have been using in that area, are ones with the faces where you are happy. You have no pain down. So you have the miserable face, and an elder will always choose the middle.

An elder will never talk about not being able to access care at end of life, because the Indian Health Services and the limited amount of funds that come in. They run out of funds halfway through the year. So elders don't seek services, because they save them for the children.

There is also the issue of some of the tribes I have worked with -- their tribal docs are doing a really good job of working with the more mainstream doctors, and some of the things that they talk about that are important -- like for example, in cancer, what is needed from the traditional docs before they can go with the more mainstream doc, if you are going to talk to them about quality of care, both of those things have to be measured.

Language is another thing.

Language is a huge issue, and the quality of care that someone who is Hmong -- I work with immigrants from 15 countries, Latin countries, and for them language is a huge issue, and it definitely impacts their care.

I could go on and on with stories, Amish. I just think that oftentimes that -- When I started in end of life in the Eighties, the thing that I always heard was back then people were trying to define a good death, and we really felt we could define this good death.

What I have learned through the years with my research is everybody defines a good death differently, and culture -- Color is not culture, but culture is really important, and I just think oftentimes that we need to look through a different set of eyes.

My last thing has to do with holistic. So I am a social worker, and I know the frustration of being a social worker in the field of palliative care and end of life, and I speak for my many, many students that have graduated and work in the field.

One of the things consistently is that sometimes there is such a medical focus, and I am not saying that the medical focus isn't important. Of course, it is important, but my research and research of others, when you look at the information that is out there, shows that psychosocial is key.

In my research I have found that pain is not even a significant variable, if you haven't addressed social support and other aspects of the psychosocial and the spiritual.

So I just think I was glad that we had Joan's work at the end, but mostly we did not have a lot in terms of the psychosocial, the spiritual, and I just think more is needed of that, if we are really going to get at a quality dying process.

DR. LUNNEY: Naomi?

MS. NAIERMAN: Understandably, we are making the first baby step in this field, and we do have some process measures that get us started with screening, assessment, discussions and so on. So I would like to prioritize those process measures to be followed by outcomes.

So having screened and assessed for pain, what are the outcomes of those? I think having done the basics for it, I don't think it will go very far, as Eduardo said, if we don't aim to complete that cycle toward outcomes.

The other thing I would stitch together is in hospice we think a lot about wasted resources, minimizing unnecessary hospitalizations, ER visits and so on. I don't think -- and yet, I don't think it is really coordinated with the other priorities that the health care system as a whole looks at.

So, for example, NQF has an NPP around use of care. I am not sure exactly what the language is. But I am not sure to what extent it has been stitched together with coordination efforts and outcome measures in the setting of end of life care.

So if you looked across the continuum, you will see that, when hospice is involved in a nursing home setting, more than likely you will see fewer unscheduled hospitalizations and emergency room visits. I think that is huge, because we are so consumed with reducing the deficit, reducing health care costs, and that is where the big money is.

So to the extent that quality can drive better resource use, I would put a high priority on that as well.

DR. LUNNEY: Solomon, I think.

DR. LIAO: I think one of the big topics in palliative medicine or palliative care that we haven't talked about is artificial hydration and nutrition.

I know that some people have looked at, for example, feeding tube placements as a quality measure, and so I am glad to see, for example, the consideration of turning off defibrillators and not giving chemotherapy in the last few weeks of life. So maybe this next time around it could fit under the same type of thinking or thought.

DR. LUNNEY: Thank you. Coming around again, Naomi, are you still --

MS. KARP: It is really just a question, because it is my first time doing this, and it seems to me that the Steering Committees are reactive in that we have to only react to the measures that are submitted, but now we are having this discussion about gaps. So I wonder what is the process for how we fill them, and is there anything we can do that is proactive?

I am not a researcher. I don't do this kind of research. So I couldn't be a measure developer, but what can we and NQF do?

MS. BOSSLEY: It is a really good question, and I think it is something that we are seeing continuing to evolve. Part of this is just putting it within the framework of this report. So these are the measures you received, and this is always one of the things most committees struggle with.

You have to work with what you have in front of you, and that often is limited within the scope of the project.

One thing that we are starting to do now, and it is throughout several of the activities that NQF has, is starting to look at the gaps. How do we begin to pull that together into reports or some way, so that it is out for others to take a look at in a more accessible, easy way.

So we are working on that now. We have actually been charged through our HHS contracts to actually develop a report on gaps at the end of this year. So the hope is that, as it becomes more publicly available out there, not only buried kind of in the report but also within a true focused document on gaps, that is part of it.

We are also trying to think, just as staff and performance measures: How do we really start feeding this information back to the measure developers, begin to know what they have in their pipeline, because often we struggle with not knowing all the groups out there that work on these issues. How do we get to them early so that they understand kind of the criteria we have, the issues that you all have faced, the other groups have faced looking at these measures, so that they can make sure they address those things.

So we are trying to figure out how to do that best, but it is a challenge. If you have any ideas in addition to what I have mentioned, please mention it, because we continue to try to figure out how to keep this moving.

DR. LUNNEY: Michael?

DR. LEPORE: An area around end of life care that we didn't address at all is really after death. I don't know to what extent it would fit in here, but treatment of the body and treatment of the family after the individual has died has not been addressed and does seem important.

DR. LUNNEY: And then ironically, I would just like to raise sort of the counter issue. I understand I was in part asked to co-chair to make sure we kept an end of life focus as well as a palliative focus, and I kind of think we drifted heavily to end of life.

I kind of look back now and say, even those of you involved in hospice, think for a minute about palliative care from initial diagnosis. Are there gaps in what we would see as good care of people who are dealing with a serious illness that may live.

I think we hit on pain, dyspnea, constipation with opioids. There is something missing from that equation. I mean, people who when they come into the health care system with a serious illness and they are struggling, let's think they are going to be alive in five years. What is the kind of care that really would mark good palliative care?

Rick?

DR. GOLDSTEIN: I don't know about in adults, but the importance of clear prognosis is -- It is probably the most determinant of how that course goes, and actually, it would be a very hard quality measure to assess, although plenty of people don't feel that they even have received prognosis -- I mean communication issues related to those patients are important.

MS. PICCHI: Actually, I wanted -- I had my card up for a little bit. I wanted to make one more comment about the previous conversation, and then address your question as well.

I was sitting here looking at the NQF preferred practices, and one of the areas that I think is really significant for families is item number 27 about educating family on a timely basis regarding the signs and symptoms of imminent death.

That has huge ramifications, obviously, for what happens with loved ones in those last days and hours, and also the whole bereavement process following and a need to really be able to attend to the psychosocial and the spiritual dimensions of the life of that family. So I would love to see a measure around that.

In answer to your question, I also would like to echo something that Tracy said earlier. That is, from the palliative care perspective, palliative care truly is an interdisciplinary art, and I don't know if we have done as much as we can and should in making sure that we have the right people that are certified and on palliative care teams available to do those initial assessments as well as follow-throughs, including not only the medical dimension but the psychosocial and the spiritual assessments and care planning that needs to be fully implemented into the continuum of care plan.

DR. LUNNEY: Thank you, Tina. Back over here, I will start, I guess, to the left, and work to the right, if that is fair.

DR. SCHROEPFER: I have such a big mouth, I forget. What Tina said -- I think one of the things that we don't -- I think that would be a good measure, too, is the training. We need more training around interdisciplinary teams and how they can truly not be hierarchical but actually work together and value each other's contribution. So that would be a wonderful outcome measure.

To your comment: My research focuses a lot on talking with people about whether they have considered hastening death and, if so, why and, if not, why not. One of the things that does come up is that, for people who have not -- so this is palliation, not talking about end of life care, but for people who are being told that they have some kind of disease, it could be -- no, it could be life limiting, but certain, it is going to be difficult to deal with. It could be that they are going to be on dialysis or whatever.

The important point is that what I found is that without support at the time of diagnosis -- so not just the conversation with the physician saying this is what you have and, you know, we are going to work with you, but more of the psychosocial, more support that way, that they often just feel like I just want to go ahead and hasten my death, get it done, because life is going to be horrible.

So having more after that conversation and making that a part of care, and how we can do that, I think, is important.

MS. NAIERMAN: I just wanted to respond to Tina. I believe that one of the FEHC measures is a conversation, communication around imminent death, and of course, integral to that is training the family about how to deal with the -- how to help and support the dying person in the home, but there is a measure within the FEHC, as I recall, hopefully, correctly. So we are measuring it, having endorsed the FEHC now.

DR. LUTZ: I think, thinking a little bit farther upstream, I think the biggest unmet need in palliative care is, if you consider from the time someone is diagnosed, they may have a chronic dying illness, four or five years, and we have discussed a lot of measures that have to do with end of life or have to do with specific topics.

There is not much out there about education of or support of caregivers. As we have 78 million Baby Boomers enter a time period when there is not enough money to put everybody in nursing homes or to have personal caregivers, it is going to be perhaps much like it was 100 years ago. It will fall to the family or someone who cares.

Those people -- We people when we are in those circumstances helping someone don't have a whole lot of resources to go to essentially to help the caregiver understand and get support. I think it is a huge issue.

MS. O'MALLEY: Oftentimes, in California hospitals in the conversation around goals of care, people who are not -- who don't speak English well, oftentimes interpreters are brought in. One of the areas that we are working on in our public hospital initiative to support palliative care is education of interpreters about how to effective in the palliative care conversation.

As part of doing that, we looked nationally for anything we could find, and we really found very little on how to help interpreters really learn a language and really be comfortable. When you think about the life of an interpreter, they go from the well baby clinic to the bedside with relatives breaking news about terminal illness.

So if there is a way to really look at the quality of the conversation for patients with whom English is not their primary language, who need interpreters, that are trained or certified -- a trained interpreter in palliative care is available for that conversation.

DR. LUNNEY: Then let's open it now to public comment and start with people in the room who might want to add to our discussion. I think I will first call on the first hand, which was behind you, Joan.

DR. LUPU: Hi. Dale Lupu. I am the consultant who works with the American Academy of Hospice and palliative Medicine on quality issues, and it has been a privilege to see the depth of the conversation today, which was really fun.

I can't bear not saying this. There is sort of a half-empty/half-full sense that I have right now. Part of me is so proud, having watched the field since its inception, that here we are having -- That we have any measures for you guys to look at is just, I think, a course of celebration, and I want to sort of say to the glass half-full, yeah. It is incredible. Then, of course, the glass half-empty is really seeing how far we have to go.

So on how far we have to go in gaps: I want to now talk -- I was kind of a staff participant with the group that is the NPCRC, the National Palliative Care Research Center, pulled together, key researchers, to bring together this package to you.

There is a letter that you received, and I just want to commend to you and maybe talk about two of those points that the researchers, in conjunction with Brookings and the Long Term Care Quality Alliance -- The two things that became evident as the measure development community was trying to pull together a comprehensive set of measures -- There were two things that became evident.

So one is, well, just conceptually -- and you guys have said this, but just conceptually I think of hospice and palliative care as having two components. There is what I would call the specialty level, which is the specialty services. So that is the patients getting hospice specialty care or palliative care/specialty care.

The primary level is all those other folks out there. You guys were talking about it. But I think it is helpful to remember that we need quality measures that address both of those slices of patients, both patient populations or both care delivery places.

When we looked at the set of measures that was coming to you that was more focused on specialty level care, so really measuring hospice and palliative care, the thing that was very apparent is that we have carved up the denominators in these small pieces. So we have the ASSESS or we have

ACOVE.

Not faulting it, but because of the NQF process, those could only come forward with the denominators that have been tested. So we are very much hoping that you will put forward in the gaps -- You guys were even trying to do it. I mean, it is very evident that this is, in some sense, a low hanging fruit, that some of these measures are now ready to be just have the denominator broadened and be able to apply either across all the specialty level care or even be pushed out farther.

So that is one gap, is let's define some common denominators, and let's get some of these measures tested, but we need the resources for the community to do that. I think that is where the report could be helpful, is saying we are really ready for this now. So that is one.

Then two is -- This is a little more subtle. That is how often we look at the other measures that are out there that are purporting to measure quality and how frequently they are completely silent about any of the hospice and palliative care domains.

Sometimes it is not intentional. Sometimes it is a lot about -- Well, for instance, in HCAHPS the way the sampling frame is set up, nobody who has died is in the sampling frame.

So there are a lot of places in which, not so much palliative care -- well, even really seriously ill folks, but especially the end of life patient population is simply excluded from the measure, both from the population, and then those concerns aren't built into those broader measures of nursing home care, of hospital care.

So that is a different way of looking at a gap, and we would suggest that that be part of the report as well. So thank you.

DR. LUNNEY: Thank you, Dale. Joan?

DR. TENO: I have three quick things to say. I actually think there is a really urgent need to define some baby steps in terms of the psychosocial outcomes and spiritual outcomes. If we don't start on that pathway, I won't have a chaplain at my IDD table.

I think we are going to have to realize that where we are going to have to start is sort of almost like a RAND-like process, Delphi panel that comes up with some things that are based on expert opinions and the best available evidence to get something out there, and we are going to need to really think about at least having our best guess of what the right thing is here, because gathering the evidence is going to take time.

My concern is, as we start sort of figuring out how to redo our health care system, we are going to go from my internship year where I used to have social workers in the hospital, where all I see is discharge planners. So I think there is some real things.

I think a real low lying fruit is grief and post-traumatic stress disorders. I think we haven't realized just sort of what you were saying before, how we manage that dying episode and what that caregiver sees and how that caregiver is educated and trained has a tremendous impact on their grief and bereavement period.

While we want to always go upstream, we need to not forget that the actual last days of life is a huge event to these caregivers, and it is a huge impact on their future life.

Then I think one of the challenges is you have people like me who are academics who tend to be the types who like granola cereal, flip-flops during the summer. We don't do really good at coming up with business plans on how to get these measures out to people and how to do maintenance.

We are more the people who do start-ups. We need to have some way of creating marriages for people who like starting things up to hand it off to other people to fully implement. As we go forward in this process, I think a lot of the people you heard present today are starter-uppers. They may not be the best person to do the maintaining.

So I think we need to think about how we can create some creative marriages to not only get the measures out there, but how to maintain them, get them used, and take it to the next level.

DR. LUNNEY: Thank you, Joan. Any other comments in the room?

Anthony, are the phone lines open, and do we have any comments from people on the phone?

OPERATOR: All lines are open.

DR. LUNNEY: Hearing no -- Oops.

DR. WHITE: Are we going to have a time for sort of process talk, like talking about how this went and what we can to improve it?

DR. LUNNEY: That would be very helpful.

DR. WHITE: Is now a good time?

MS. BOSSLEY: I think -- Yes. June agrees, yes, now it is time. I have been at least hearing and participating in quite a few side conversations, and I would think it would be very important to put it out here in the center of the room.

DR. WHITE: Okay. It seems like a lot of our conversations hinged on just definitional uncertainty today. We had lots of questions about what is what. What is reliability; what is validity.

I think I talked a little bit about a cheat sheet yesterday. It would be great to see for each of the key criteria that we are supposed to be using to evaluate this a very pithy summary of what is reliability. What are you looking for. What are examples of good evidence, moderate evidence, poor evidence.

I know that we got this very long packet, but it was -- It didn't sort of fulfill the brevity need, the sort of pithy -- Give us the summary that we need to sort things out.

MS. BOSSLEY: We often struggle with that. I will admit it. Any thoughts you may have on how to go about doing that, because you are the end users -- We are in the weeds on this all the time. So I think sometimes it is harder for us to get out of it and think, if I were someone new to this, what would be useful.

So any thoughts you may have are welcome, because we know that this is an evolving process.

DR. WHITE: Two quick ones. Some

of us are, and some of us aren't, researchers, and even among those of us who are researchers, not all of us have really thought about reliability and validity, and yet these are central concepts. So maybe just a little pithy primers on what is reliability, what is validity, what are the different kinds of validity, and what is most important.

DR. CASARETT: I actually had a thought along those lines. Although I really appreciate the free flowing nature of these discussions with back and forth, bouncing around, part of me, though, would welcome more NIH-like study section structure in which you walk through set criteria and say, in general the group thought that this met these criteria, validity, reliability, because of this, and maybe even, going out on a limb, preliminary reviews, so we have in front of us as we are voting a reminder that the reliability was thought to be good because of this correlation coefficient, based on this number of samples.

I am not sure if that is -- That may be too much structure, but I think a little bit more structure might help.

MS. BOSSLEY: No, it is a good point. The other thing I will mention to you is we are looking at having outside consultants provide an analysis of the reliability and the validity testing, moving forward.

So, actually, every measure would have -- and we have done it in certain projects. We had an outcomes project last year, and we have a resource use project this year that we have actually provided that type of information, so that you have an outside expert taking a look at that.

So it sounds like moving in that direction might be a good thing on all projects, which is, I think, where we are headed.

DR. LUNNEY: Thank you, David.

Robert?

DR. FINE: My question, and I had asked Karen early on when I was about to veto all of these things in previous review, and said why am I going to D.C. if I am going to vote against all of them.

There is this whole section on staff notes. It wasn't completed, and the answer, I think, Caren gave is, well, we are not sure we had time, but I am curious what the purpose of that is.

I could see it being very useful, that if there was someone with expertise in looking at these -- I kind of consider myself a reasonable content expert after 25 years of working at the interface between life and death, pretty good on that, but I am not a statistician.

I don't know those other things, and because I have not been a researcher -- I have been a clinician and really tried to implement what researchers do out in the private sector where everything has a really strong financial model. It doesn't happen in my sector.

I wish those staff notes had been filled in. I think that might have helped me, given my deficits in some of the areas of statistics and all, if that is someplace where they would have put that in.

DR. LUNNEY: Naomi, on the left.

MS. NAIERMAN: Very quickly, the slides that you had did give a very brief outline of each of the measures, and I found June's pithy words about each one of those technical language descriptions even more helpful.

So I think, if you put even a sentence after you have put the criteria in one or two words, that may be a start toward a brief cheat sheet.

DR. LUNNEY: Naomi, on the right.

MS. KARP: I think most of my points have been made, but I am going to just chime in with the -- I know a lot of substance about this, but I am not a statistician. I don't know that language.

I think you perhaps erred on the side of you gave us so many tables and charts and so much material that, when I started to do the process, I was just overwhelmed. You know, I had so many documents open on my computer that I didn't know how to toggle back and forth between them.

So I guess I will echo what everyone else said: The cheat sheet, that plain spoken sentence or two about what does this really mean. It would have helped me a lot get over the psychological hurdle of, oh, my god, I don't know what kappa means. So how can I do this?

DR. LUNNEY: Thank you. Eduardo.

DR. BRUERA: I personally had much -- I had limited problems with the reliability, validation and all that, and I think our voting suggested that that generally was okay.

The problem is when you have outcomes that are not measured by psychometric analysis, like getting chemo before you die, going to the ICU, disconnecting your defibrillator. I mean, you don't need reliability, validity to turn off or to put a magnet on a machine. It is just something that you do or you don't do.

So the results are looked in different ways. The outcomes are measured in different ways than reliability, validity of psychometrics. I find that the analysis of the literature for that is way more complex, because you have multiple sources.

I would have loved to see that we got a little bit more kind of bouncing back and forth with the instrument developers about what is missing there, and I don't know if the peer review process before presentation to make sure that the methodology followed this more solid would help the process of this meeting coming together, because in some cases what was missing was a lot of information that I had to go and do some PubMed searches and look around myself, and the research was done by other groups, and sometimes with totally different questions.

so it gave me satisfaction that I understood the problem, but you know, these are volunteer groups submitting volunteer information to other volunteer groups, and perhaps getting some peer review where it has been submitted would help us an awful lot in the fact that when outcomes are used rather than the pure methodology of the instrument development, you have to go a little bit deeper, and that takes hours of work.

DR. LUNNEY: Kate. Sorry.

MS. KALEN: I guess part of what I reflected on over the course of the two days was the bouncing a little bit back and forth when we were reviewing the measures between how strict did we have to be to adhere to exactly what was presented to us in the materials versus the expert opinion that we have in the room, where we might not be specifically citing other studies that would support a deficit where there was a known deficit in the materials presented that someone here might have been able to fill in the pieces.

That kind of shifted back and forth across the two days, and I think created some uncertainty as to, for the reviewers, how much latitude did we have, if there was no information presented for disparities or something and the developer doesn't address it in the conversation, and yet we know e can connect the dots to know that particular measure could be used to -- you could use it and figure out the disparity question later.

I guess there wasn't enough clarification for me on that.

DR. LUNNEY: Thank you, Kate. Sorry, I didn't really skip you. I just saw hers first.

MS. O'MALLEY: A couple of concrete suggestions. I think this is a process that would be enhanced by some kind of a checklist, particularly for people who have never done it before. It would have really helped me organize my thinking. Like Naomi, I was like where do I start on something like this.

Something else, if you have the resources, that might be helpful for people who are not health services researchers is to have a fast paced tutorial that you could just go through on the web, that would be like this is what this is, this is what we are looking for, and within that to embed some examples of best practices: This is the best answer we have seen to this question, also paired with this is the worst answer we have seen to this question. So you know if you are voting somewhere in the middle, you know what your conscience is telling you to do actually has some face validity to it.

Then I think also it would have been really helpful -- I am thinking about all the ASCO back and forth that we had yesterday, and I as part of the I team looking at this, realizing that I thought a lot of it was insufficient.

It might have been helpful to have it stopped at the gates; you know, to have a staffperson say this is a maintenance measure that is coming back; we have a new set of criteria, and send it back to the developer and say we don't think this is going to get very far in the process unless you consider X, Y and Z might be a kind thing to do, rather than bring it forward and then have it go back anyway.

You on the inside, knowing much better than we do what your processes are, now in hindsight, I am saying, gosh, why did we even spend so much time. It probably would have been kinder to send it back to the developer and say we need a little bit more substance here before we feel it is going to meet the new criteria.

So a couple of suggestions.

DR. LUNNEY: Solomon.

DR. LIAO: I also have a couple of back and forth suggestions. So along that same line, if I can ask that when we do the first measure that we do it with a developer that is actually in the room or at least available -- an easy one, an old one that you are virtually guaranteed consensus and all that.

MS. BOSSLEY: We typically do do that. For whatever reason, it didn't happen this time. So I apologize.

DR. LIAO: Then secondly,. in terms of our orientation, I would have found it very helpful that, instead of having a staff member walk us through -- or maybe in addition to the staff member walking us through the review process, to actually have a prior committee member who has actually reviewed these show us their thinking process, going through it.

DR. LUNNEY: And I am particularly guilty of having changed considerably my review in this room from my review ahead of time, because I was under the understanding that, if it wasn't on the paper in front of me, that I didn't have to presume that it existed somewhere and go looking for it.

I think that is a real tough issue. There were some real differences here in how much trouble people went to to complete their application, and I, with my NIH background, get annoyed with people who don't want to take the time to provide that information for me, but expect me to already know what they know.

So I wish that could be evened out a little.

Do you have another?

DR. WHITE: Just a quick follow-up

on that point exactly. It is the same point, for sure. Can I just round up that one.

I came to this late. So I wasn't part of the early calls, and maybe you guys did this, but it would be one thing to consider about how to get us all on the same page about how to do the evaluations is to devote 15 minutes and say, you guys are going to in the next month be asked to do all these evaluations; here is -- we will take you through one.

DR. LUNNEY: Over here. Go ahead.

DR. SCHROEPFER: I have a question, and maybe I missed it. When we entered on Survey Monkey, was there a way to save part-way through? No. I had an emergency with a student, and I had to go, and I had put half of my stuff in, and then my computer restarted with an update, and I lost everything, and I had to start all over again.

So I think having that Save button would be like the most wonderful thing.

The other thing, I think, yesterday I felt like when we came in that it was kind of the agenda was set out for us to start right away with the measures, but some of what I think happened, too, was just the need to kind of debrief for a moment, just to kind of talk about some of the ground rules and how we were going to proceed.

So I think that some of that, sticking with that first measure was actually us seeking, whether we were conscious of it or not, really trying to figure out what did that mean. What was it going to look like.

DR. LUNNEY: Any other tents?

DR. CASARETT: If there is a way to do that in advance at future meetings, that would be great, because time when we are actually here is really valuable, and that is something that could be -- at least be done on a call.

DR. LUNNEY: I don't see any sideways tents. Then I think it is time for you guys to tell us --

MS. BOSSLEY: Thank you for these comments on the process, because we appreciate -- You are the first ones actually going through the new criteria, which was why you kind of see Helen and Caren and myself going, is this working the right way or not; because, unfortunately, you were the guinea pigs.

If you, over the next few days, reflect back and have any other suggestions, please tell us. We have another committee on renal disease that will meet in August. So we have got to huddle now and figure out how do we keep improving the process, trying to reflect what our business is, to make sure that we do it with our committees as well.

So thank you for everything. For next steps, I think, Caren, do you want me to go over something, or do you? Okay, I will Caren go over.

DR. GINSBERG: The list -- to-do list from yesterday was to get back to Craig

Earle about the one, two, three, four, five, six, seven measures we need input on.

We have to talk about harmonizing pain measures and finding a common denominator, a numerator for different denominators.

We tabled the spirituality measure, and we need more information from the developers on that.

We have to talk about the treatment preferences; measures. There are three that have to be harmonized, two from this meeting and one, an existing measure 326 on care preferences.

Then everything else you just said to us in the past few minutes go on the to-do list, too. We have to work through that as well. Thank you for those comments.

Is there anything else that we need to get back to you on immediately about the measures?

We need to reschedule. Sean has

another commitment. So we will have to reschedule. We were hoping to do it in August.

MS. TIGHE: We are looking to schedule it about three weeks from now. So I will be in touch to figure out a date that works.

MS. BOSSLEY: And keep in mind, I have the feeling, because we have seven measures that we tabled, we will probably have to have either multiple email exchanges or potentially a second call.

So we will just have to play that be ear and see how we move forward, but I think, given the amount of work that you may have in front of you still, you may have two calls coming.

DR. GINSBERG: If you could get any questions you have for Craig Earle to me or Lindsey or Heidi within the next couple of days, certainly by Tuesday -- we are trying to set up a call now. So if you have questions you need answered, just get them to us.

MS. TIGHE: Actually, if you could do it before the weekend, one of the potential times to speak with him is on Monday afterwards.

DR. GINSBERG: All right, tomorrow. I was trying to be nice.

(Whereupon, the foregoing matter went off the record at 2:52 p.m.)

CERTIFICATE

This is to certify that the foregoing transcript

In the matter of: Palliative Care

Before: NQF

Date: 07-21-11

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R Gross

Court Reporter