NATIONAL QUALITY FORUM

Moderator: Benita Kornegay Henry
June 20, 2019
4:35 pm CT

(Sam Stopey): Hello and welcome everyone. I'm (Sam Stopey) with the National Quality Forum and you are attending the Patient Experience of Functions Spring 2019, Measure Review Cycle Standing Committee Meeting.

I just wanted to give a brief word of welcome. Delighted that you're all here on what's undoubtedly going to be a busy day for all of us. We're going to try to get through as much of this material as possible, but -- as you know -- for the portions we're not able to complete we do have follow up meetings and we'll be certainly using at least one of them.

So -- excuse me -- we're going to move directly into our agenda. But our first order of business is around disclosures. So we'll allow a brief welcome from our co-chairs who will invite you to this disclose at that time.

And then we'll turn it over to our Senior Vice Present of Quality Measurement -- (Alisia Mendally) -- who's going to finish the disclosures.

Lee Partridge: Good morning everyone. I'm Lee Partridge. It's very nice to see all of you here again after what's been an awfully long time.

As (Sam) already has said we are going to be brief in everything we try to do today. But I - and I - the co-chairs and the staff have talked a fair bit about length this agenda.

We are going to try very hard to be disciplined and careful of that keeping ourselves more or less on time. But I don't want that to interfere with the richness of our discussion.

So (Chris) and I are going to be martinets. But -- still -- if you feel at any point that we're trying to shut you down just override us and keep going.

As we go around the room also it would help us to know if any of you have to leave before 4:30 - well tell us if you have to leave before five, because it will help us figure out what we have to absolutely try to cover today and what perhaps can move over to 2:00 on Tuesday for our next meeting over the phone.

(Chris Stilla): And good morning everybody. (Chris Stilla). Excuse me. Greetings from Colorado where it's nice and green. Come visit.

I think we said basically everything, you know. I think there's a lot of important stuff that you all have in your heads, keep you briefed and synced. We will get out of here having done a whole lot of work. And I have nothing to disclose.

Woman: So good morning everyone. My name is (Alisia Mendally). I'm Senior Vice President of Quality Measurement at the National Quality Forum.

I wanted to welcome you -- again -- on behalf of NQF and thank you so much for being on the committee.

And (Sam) and Lee and (Chris) has said we are going to combine introductions of the committee with disclosures of interests.

When you were first nominated to the committee you answered a number of questions about activities that are relevant to the work in front of you that were either paid or unpaid. And so today what we want you to do is verbally disclose any - anything that you put on that form that's relevant to the work in front of you.

Just a few reminders. You sit on the committee as an individual. You do not represent anyone who may have nominated you for the committee or your employer.

We are interested in both paid and unpaid activities as they're relevant to the work in front of you.

And probably the most important reminder is just because you disclose does not mean you have a conflict of interest. We go through this process in the interest of openness and transparency.

And so I will start with your co-chair. So I'll start with Lee first. I'm going to ask you, Lee, introduce yourself again. Let us know who you're with and let us know if you have anything to disclose.

(Unintelligible) (Chris). Then we'll go around the table clockwise starting with my left, (Don).

Lee Partridge: Lee Partridge. I'm currently a half-time volunteer with the child health team at the United Hospital Fund in New York City, which is a 140 year old (unintelligible). And I have nothing disclose.

(Chris Stilla): And (Chris Stilla). I'm an Academic General Pediatrician at the University of Colorado and Children's Hospital in Colorado and nothing to disclose.

(Alisia Mendally): Thank you. (Don)?

(Don Casey): Thank you. (Don Casey). I'm an internist -- now President, rather than President elect -- of the American College of Medical Quality faculty member of Jefferson College of Population Health Rush and Minnesota Institute for healthcare and (unintelligible).

My only disclosure is that I was the initial chair of the committee that approved the first patient experience measures back in 2006. That's probably not a conflict, but it is a disclosure.

(Alisia Mendally): Thank you.

(Gary O'Mally): Hi. (Gary O'Mally), Geriatrician (unintelligible) Hospital, partner (unintelligible) in Boston, Harvard Medical School and I have no conflicts.

(Alisia Mendally): Thank you.

(Lisa Suiter): (Lisa Suiter). I'm a Reumotologist and internist at Yale.

My conflicts are that I volunteer for the American College of Rheumatology as their quality measures sub-committee chair. I also work and have salary support for (CMS) contracts at the center for outcomes research and

evaluation at Yale as I'm a director there. None of these of (ACR)s or cores measures are in front of the committee today.

I have with (unintelligible) (Cleary) who's at Yale and who's involved in the (CAP)s measures been to a meeting of (CAP)s in the last fall where I participated in a - in discussion groups, but I don't have any influence over the measures.

(Alisia Mendally): Okay.

(Deb Soliva): Good morning. I'm (Deb Soliva). I'm a Geriatrician and health services researcher. I direct the UCLA (unintelligible) center for (unintelligible) research.

I'm also in the veteran's administration and I work also at the (RAN) corporation as a Senior Natural Scientist.

I have - I'm the immediate path or chair for the American Geriatric Society. And I do not have any conflicts with any of these measures. Although I have received multiple grants and funding from several of the funders of several of these measures. But I discussed it with the counsel and it was not deemed to be a conflict.

(Betsy Rebek): (Betsy Rebek) from health (unintelligible) in Minnesota. I'm an internist. I also am the chair right now (unintelligible) measurement who has used some of these measures in the past and then within our organization health partners we have also used some of the (CAP)s measures. But I don't know if that would be a conflict for what we're doing today.

(Lisa Maurice): I'm (Lisa Maurice). I'm the Executive Director of (unintelligible) advancing patient (unintelligible). I represent patients and caregivers.

And my organization is a sub-contractor in patient and family engagement to (CMS) on the partnership for patient (PIN) project and the transforming clinical practice initiative project. We don't do the work with measures.

I am also on the clinical practices committee for (M22A), which uses measures, but I'm not involved with any measure and usage, so - anyway.

(Alisia Mendally): Thank you.

(Dawn Hong): Good morning. My name is (Dawn Hong). I'm a nurse by training. My role is within the John Hopkins Health System. I oversee transitions of patients going home to home based services. And then as it relates to patient experience I oversee the (HH) (CAP) for the home care group.

I don't believe I have any conflict of interest, but I would disclose I work closely with (HH) (CAP)s day to day with analyzing results and educating our teams on the results.

And the only scheduling conflict -- I just wanted to bring up -- I don't have any conflicts today, but on Tuesday I cannot be on the clock in two to three. So…

(Sharon Cross): Hi. My name is (Sharon Cross). I am a Oncology Social Worker by training. I work at the Ohio State University Medical Center.

And my day to day role is patient and family engagement. I work with our vast network of patient/family advisory councils that we have.

I don't think I have any conflicts of interest to disclose.

(Ann Monroe): Good morning everyone. I'm (Ann Monroe). And I represent - I hold the consumer seat here. I've been in Philanthropy for about 20 years focusing on people and their issues and how to improve healthcare for them.

I don't think it's a conflict, but I do share the oversight panel for the $6 billion Medicaid waiver in New York State and we look a lot at how people are performing against the measures, which I don't think it's a conflict.

I was chair of (CSAC) here when we looked at a lot of these measures the first time out, right, Lee?

Lee Partridge: Yes.

(Ann Monroe): And some of them look exactly like they used to look. It's kind of interesting.

But - so I don't believe that I have any direct conflicts.

I will need to leave about 4:30 or 4:45. But hopefully that will be okay.

(Brian Limburg): Good morning. My name is (Brian Limburg). I am the Executive Director for the Consumer Collision for Quality Healthcare and also a Public Policy Advisor to a number of 18 groups here in D.C.

I have no disclosures. I do have to leave at 3:30 though.

(Alisia Mendally): Okay.

(Ryan Collar): Good morning. I'm (Ryan Collar). I'm a Academic General Pediatrician at the University of Wisconsin, American Family Children's Hospital. Focus on hospital medicine and conflicts care and do health services research as well. No conflicts.

(Alisia Mendally): Thank you.

(Steven Hoy): Good morning. I'm (Steven Hoy). I'm also sitting as a patient family representative. I work as a Chief Operating Officer at (PFCC) Partners. I have no disclosures and I'll be here all day.

(Ellen Sholts): Hi. I'm (Ellen Sholts). I'm a Senior Researcher at the American Institute for Research. I have worked within the past five years on a (CAP)s measure, so according to NQF policy I'm refusing myself from review of all the (CAP)s measures this cycle.

(Sherry Ericson): Hi. I'm (Sherry Ericson), Vice President for Governmental Affairs in Medical Practice at the American College and positions, which represent internal medicines position to use all of these measures essentially or most of them.

I do - in my role there I do write most of our public policy related to the use of performance measures for public reporting and payment policy. But I don't have any funding from any measure developers.

I also am on the steering committee for the (CQMC), the core measures collaborative. I don't believe that's a conflict, but just want to disclose that for the group here.

And in my role at ACP I have also worked with (NCQA) to develop an alternative payment model that is going through the (P-check) process. It's

specific to internal medicine specialists largely and their work with primary care and care coordination in that space.

But I don't have any funding from (NCQA). You know, just with that program. If it were to go through we'd use probably many of these measures.

(Peter Thomas): Morning. I'm (Peter Thomas) with the Power of Law Firm here in town. I represent a fair amount of rehabilitation, disability and healthcare clients and work on quality issues and measures all the time, but really from an advocacy and educational, kind of informing clients kind of perspective. I haven't worked on any of the measures before us today in any specific way. And I have no conflicts of interest. Thank you.

(Linda Maloum): Hi. I'm (Linda Maloum). My training is in Forensic Psych and health law. And I've worked in - as - I'm currently working as a risk manager. But I've also worked as a director for patient experience for a health care system in western Pennsylvania. And I worked for (Spalding) at - in the post-acute sector as well. I don't have any conflicts and I believe I need to leave around 4:15.

(Brenda Lee): Good morning. My name is (Brenda Lee). And I am President of (Leethum) Associates. I was formally a Senior Study Director of (Wesbit). And (unintelligible) I did work on (CAP)s measures and had several (CAP)s projects of which I was not a part of.

I also in my current capacity serve and preserve (unintelligible) factor of the Pathways Community Hub institute, Pathway Certification Program.

And just by way of background my background was in health policy and I'm pursuing studies in bioethics.

(Sherry Cathlin): I'm (Sherry Cathlin), University of California Irvine. I'm a Psychometric Ian by training, which my grandchildren pronounce psycho magician. I'm a…

Man: I like that.

(Sherry Cathlin): I'm the system (unintelligible) for healthcare measurement and evaluation.

I have a bunch of disclosures. I don't think any of them are conflicts. But I serve on the NQF (unintelligible) panel. I also serve on the NQF all (unintelligible) admissions panel.

I'm on the technical advisory panel for physician compare that use the (CG) (CAP). I'm on the Yale team (NIP)s in-patient measures group. I'm also a consultant of (MGH)s shared decision-making team.

And I just completed a grant to redevelop my measure of children's self-reported health state using animated touchscreen-based technology for children ages 4 to 12. But that grant is now over.

I have another grant in, but children are not -- as far as I know -- I can't remember all the measures, but I don't think that's one of them.

Woman: Well that's good.

(Chris Stezzy): (Chris Stezzy) (unintelligible) for healthcare quality and performance measures.

I'll probably disclose that I try to support that NQF's incubator wherever I can to get to measures that matter or patients and providers.

No conflicts.

(Alisia Mendally): So thank - I thank everyone and, (Ellen Sholts), thank you for disclosing your conflicts and -- as she mentioned -- she will not be reviewing those measures or participating in this session of the (CAP)s measures, but she'll be in the room. You don't have to leave the room. And, (Lisa), I see that you have a question.

(Lisa Suiter): I did not.

(Alisia Mendally): Oh.

(Lisa Suiter): Let you know that I need to leave at 4:00. I apologize.

(Alisia Mendally): Okay. Great.

So before I turn over the meeting to (Susan) I just wanted to remind you if at any time you remember that you have a conflict we want you to speak up. You can do so in real time or you can come to any one of us in the front, your co-chairs or the (unintelligible) staff.

And -- likewise -- if you believe that any one of your colleagues is acting in a biased manner we want you to speak up. So thank you.

(Sam Stopey): Very good. Thank you everybody. We're going to go ahead and move into the slide portion of our agenda. So believe it or not we are ahead of schedule.

Man: Right.

(Sam Stopey): So it feels good. So let's try to keep it up. So - and - but this portion we're just going to do a walk-through on our project introduction and full review of the evaluation process. I'll hand it over to our project's - Senior Project Manager, (Susan Sparcet) for this portion.

(Susan Sparcet): Great. Thanks, (Sam). Good morning everyone. Wonderful to see you all in person again. It's been a while. So I'm going to go through these very quickly, because I know you all have been through a few evaluation cycles now and are quite familiar with the rules of the road here.

So just briefly we ask you to evaluate all the measures other than those that you are refused from of course against all of the criteria. Discuss each criteria and vote on that criteria and then we move onto the next one.

We'll be making recommendations this morning to the NQF membership for endorsement. And then we also ask you generally to oversee the patient experience public and function portfolio measures.

Next slide. Our co-chairs will be here to facilitate the meeting ably and we appreciate their role in helping us stay on track today.

Next slide. And - next slide. And you can skip that one. And just a reminder that NQF staff are here to assist the committee in making their recommendations and providing any materials that you may need as you review and then we'll be writing everything out.

Next slide. All of the measures under discussion today did go through the methods panel. Because they are all complex measures they were all - the methods panel believes that all of them are reliable and valid. And you will

have the option to either accept the methods panel recommendation for almost all of the measures or you can discuss and make your own decision on those.

We do ask you to especially consider questions around measure exclusions, measure the populations included and excluded from a measure, as well as any questions on the risk adjustment model, because the methods panel does not have the clinical and topic familiarity with each of these measures that we would need - you - that we need you to have.

Please let us know if you have any questions as we're going through about the methods panel and we can do our best to answer those. We have included their evaluation results in your packet.

Next slide. And -- again -- just our general ground rules. Just be sure evaluation recommendations on what's before you and whatever information the developers provide verbally. Stay here and engaged as much as possible and I - we look forward to all of your contributions today.

And next slide. I do want to spend a minute on the measure of discussion process. We will have our measure developer colleagues join us either here at the table or via the phone and we've given them each a couple of minutes to introduce their measure, two to three minutes each, except in a few places where we have multiple measures from one developer or there's been a request for additional information, we've given them a bit more time. But we do try to keep those pretty brief.

And then we will ask a lead discussions to open the committee discussion by flagging any issues that came up in the committee free evaluation comments, any issues that you saw in your deep dive in the measures. And then emphasizing any areas that you think the committee should discuss again

starting with importance and then we'll ask you to start off the discussion for each criteria.

And the developers will be here available to respond to your questions as needed. And then -- of course -- we'll ask you to vote after you finish each of your discussions.

Next slide. All right. I'm going to - next slide. Just spin through these very quickly. Next slide.

As a reminder we will be looking first at importance, which is evidence and then gap, two separate discussions, two separate votes. Then reliability and validity. Again, you can either chose to take the methods panel recommendation in all except one case or you can discuss that and vote and make your own decision.

Feasibility, it comes next and then Usability is a vote and then use is a vote. As a reminder evidence and gap and reliability and validity are must pass for all measures. Usability is also a path for maintenance measures.

We do have competing measures before us today, but we ask you to hold all discussion related to competing measures until the afternoon session when we have that specifically on the agenda, because we need to have discussed both measures in the pairs prior to having the competing measure conversation.

And, (Jordan), if you could just jump ahead unless anybody has any questions on the criteria. Could you jump ahead just to the competing slide. I'll just pause here and see if anybody has any questions while (Jordan)'s moving ahead.

We'll go into a bit more detail about the competing measures right before we begin the discussion. We'll talk about the criteria and NQF's algorithm for that. But just to remind you a measure is considered competing. If it has the same team target population and the same measure focus.

So - just back up a couple. No problem.

And - okay. We're good. I think. So we'll talk more about competing measure criteria later this afternoon.

Next slide. We are looking for measures based on (ICD) pen-coded data, which these all are.

Next slide. And then - so as you know we're going to get through as much as we can today. We will be discussing some measures we expect next week on the calls on June 25. Actually we know - we had a last minute change and we will be discussing each (CAP) on June 25 and I think possibly (unintelligible) (CAP)s as well due to developer - something came up.

And so we'll be making recommendations. Staff will write up the report. We'll put that out for comment and then we'll bring you back together in September to discuss the comments that were received. Make it - if you didn't achieve consensus on any measures at this time you will be asked to try again at that meeting and these will go to (CSAC) in October for their endorsement decision and then out to appeals.

Next slide. Next slide. I want to just briefly talk about the voting process. We have a new - since you were last here we have a new voting system. (Jordan) sent the link out to the committee members via the calendar invite. So please open your calendar invite to get that link. That will allow you to vote. And

we'll ask you at - we'll let you know, we'll open up the vote, have you vote and then we'll close it and announce that result. So if you have any trouble accessing that link let (Jordan) or (Navia) know and they can help you get on that voting poll. And we've had much better luck, so hopefully that will go smoothly today.

In terms of achieving consensus we do - we have a (unintelligible) quorum of the committee. We have 21 folks in the room I believe. And that's 66% of the committee, so if we lose seven of you then we have lost quorum and we won't continue to vote, although we can discuss.

Pass or a recommended is greater than 60% not inclusive of 60 voting yes or high plus moderate. Consensus not reached is 40 to 60% of the committee voting either yes or high or moderate and that's inclusive of 40 or 60 and at that point if we don't achieve consensus on a must pass criteria the committee continues their conversation, but does not vote on the overall recommendation for endorsement. We'll hold that discussion at the first comment call.

And then does not pass is less than 40% yes votes of the quorum. And -- again -- that does not include 40.

Next slide. Questions? Either process or a criteria.

Man: Yes.

Man: Not at all.

Man: (Unintelligible) question I just wanted to ask. I've been doing this for a fair amount of time at this point and like the comment I heard earlier a lot of these measures seem familiar. And I'm just wondering in terms of maintenance

measures has NQF ever pulled back an endorsement of a - that was of a measure that was up for maintenance? And if that's the case what are the - what would constitute a reason to pull back an endorsement of a measure that's up for maintenance review?

(Susan Sparcet): Well our criteria have changed especially for those of you who have been around our table for many years. You'll know that things have gotten more stringent and we ask for more.

So something would lose endorsement if it no longer reached our criteria. Or if there was a better measure, competing measure came along and the - a committee elected to choose that one. So yes, endorsed measures can and do lose endorsement.

(Alisia Mendally): And I would also mention to (Susan)'s point if the performance gap is narrowing and we are making progress often times that is a queue to committees that we probably don't need this measure anymore and so they may not renew maintenance of endorsement.

Man: Thank you.

Man: Just the other thing that I've seen in some of the preliminary stuff is use and usability is a lot more important than say the psychometrics. And so that we still have to focus on.

(Susan Sparcet): We do actually slightly change the emphasis on the criteria. Then we have a couple slides. (Jordan), can you jump back to the importance, the table - we have a couple slides on how the criteria changes, but it's - we look for new - if there's any new evidence and the evidence has changed we would want to know and I think actually all of these measures have updated evidence. But if

the evidence hasn't changed and then the committee might not need to look at it, but you might need to know that, you know -- hey -- maybe recommendations have changed and that process is no longer recommended.

Again, we look at it - for maintenance measures gap in care and then can you jump to the next criteria emphasis slide? Next. Next. Next. Next. Next. It's in here somewhere. Keep going. Here it is.

Again, for maintenance measures we would have decreased emphasis on - maybe on the testing, but some of these measures would not have been adjusted for social risk, the last time you looked at them. So now we would look at that and that might influence your decision.

Woman: Just to clarify back to the evidence. So with a maintenance measure would we be looking for the developers to say clearly over this period of time, "This measure has had this positive impact on this group of patients."

Man: (Unintelligible).

(Susan Sparcet): That would be the hope, you know, I mean…

Woman: Okay. (Unintelligible)…

(Susan Sparcet): …sometimes…

Woman: …you could be clearly looking for evidence of the time and money spent on this has had an impact on this outcome?

(Susan Sparcet): You would hope. I mean I think sometimes there is low improvement in care for reasons, you know, sometime - we also see this in the - later in the criteria

where we look for performance over time. I mean I think sometimes, you know, we see something like, "Well this measure -- when we first submitted it -- only 50, you know, like, 500 providers were using it and now we have 5000 providers using it." And so it's - the pool may be a different pool. It's - maybe it's different patients, maybe it's different providers.

You know, so you have to take all of that into account.

Woman: All right. How does the new voting system handle conflicts or refusals? Is there a spot in it where you just (unintelligible)…

(Susan Sparcet): Just the stone vote. Just…

Woman: Hands off?

(Susan Sparcet): Yes. So we'll just take you out of the denominator for the vote, so…

Woman: Okay.

(Susan Sparcet): …instead of looking for 21 votes we would look for 20…

Woman: Okay.

(Susan Sparcet): …when you're refused.

Woman: And then…

(Susan Sparcet): (Unintelligible)…

Woman: …just one thing about following of the maintenance measures, because when the scientific methods panel sees either a change in how a measure's being used with the - level of analysis and wherever it's expected of that is going to manifest itself in the testing presentation of the -- in what's the -- kind of testing that they did.

But also NQF criteria are evolving over time.

Man: Yes.

Woman: And so just because a measure was approved three years ago, three years ago things have changed (unintelligible). So, you know, it's kind of - that's a tough one and from perspective, because when we've seen things come through now they've, you know, some of these maintenance measures have not updated into the kind of standards that are now being used and that is a little bit disconcerting.

(Susan Sparcet): Yes.

Man: (Chris) and then (Deb).

Man: Yes. A number of the submissions that were a application indicating new evidence and I was struggling to try to find out what that new evidence was other than the submission of the testing or the experience data. So the experience data measure constitutes new evidence? Do you…

Woman: (Unintelligible) yes.

(Susan Sparcet): Yes.

Man: Yes.

(Susan Sparcet): But we do have a preference for a systematic review of the evidence that has, you know, that meets the quality, quantity and consistency.

Man: Sure.

(Susan Sparcet): If there is a maintenance we want to see directionally that it is the evidence that's pointing in the same…

Man: (Unintelligible).

(Susan Sparcet): …direction. But that is - we have a, you know, you rate the measure to the degree that it needs the certain criteria, some criteria.

Man: Yes. The reason I ask is that I just (unintelligible) my bias is when I think of evidence (unintelligible) things of that nature. I didn't see any new ones other than the feedback on the measure. And that's good. That's fine.

(Deb Soliva): I thought in reading some of my materials that I read that the maintenance measure was not being held to the same reliability requirement as new measures. Is that not correct?

If that - you know, that we have - we changed and have somewhat (unintelligible) evolved in our definition of reliability and that for the maintenance measures we were not necessarily applying that. That was in the materials for one of my measures that added to the maintenance measure. Okay.

(Susan Sparcet): No.

(Deb Soliva): Okay. All right.

Man: (Unintelligible).

(Susan Sparcet): If you find that let us know, so we can…

Man: (Unintelligible).

(Susan Sparcet): …(unintelligible).

(Deb Soliva): Yes. Okay.

Man: It would be clear. The answer is they're the same as (unintelligible)…

(Susan Sparcet): Yes.

Man: Okay. Thank you.

(Susan Sparcet): Yes. There was a time…

Man: (Deb)…

(Susan Sparcet): …when…

Man: …you're not alone. I think I read somewhere in just the volume of documents that with maintenance measures validity and reliability are not as emphasized as much. And other specters -- like you said -- gap and other things are usability are a bit more relevant. (Unintelligible) I think I read that right.

(Susan Sparcet): They aren't as emphasized as much if they're up to the current standards.

I wonder if what you're thinking of is - we do now require that all meet measures have empirical validity testing. But - and (unintelligible) we used to allow measures to have (unintelligible) validity only - sorry. We allow all (unintelligible) require all maintenance measures to have empirical validity testing, not just (unintelligible) validity.

But we will occasionally get a measure where they were - there was very good reasons why they were unable to do the empirical testing and they are allowed to submit a justification of why they only have (unintelligible) validity and the committee can discuss them, decide whether or not to accept that.

I don't believe we have that in this set of measures though. But that may be what you're thinking of.

(Deb Soliva): It was in the 2635 language for 2635.

(Susan Sparcet): Okay.

(Deb Soliva): But -- of course…

(Susan Sparcet): That's really (unintelligible) needle in a haystack, because that was a 140 pages of material. I'll take a look.

Man: (Unintelligible).

(Sam Stopey): Any other questions at this point?

Woman: Just quickly, in one of the measures on the questions for the committee it says the scientific method panel is satisfied with the reliability. Does the committee think there is a need to vote on reliability? Don't we vote anyway?

(Susan Sparcet): You would vote to peek the committee - the methods panel recommendation.

Woman: Okay. But we're going to vote on every one of these?

(Susan Sparcet): Yes.

Woman: Yes. Okay.

(Susan Sparcet): Yes.

Woman: The scientific methods panel sometimes disagrees with itself. You know, we have whole throated discussions about some of the issues at hand. And so sometimes you'll see -- even though -- it - the recommendation says pass, some members felt there was either insufficient or low reliability and there are reasons for that.

So maybe, you know, just sort of accepting the recommendation may not be enough if the committee feels like there should be a further discussion about what went on.

Man: And the other caveat to that is that the expectation is for the scientific methods panel is to serve in an advisory capacity to the committee in putting forward the recommendations from a mythological standpoint.

However, there's always going to be some sort of the clinical or other component where we're going to rely on the expertise of this finding to

analyze as (Sherry) said. The methods in how the committee -- or sorry -- the methods panel actually voted as well as to consider other factors that may influence the reliability or validity of the measure.

So you may elect to verbally -- we'll just put it out there verbally -- would you like to report or does anyone wish to decent from that? And then in which case we'll move directly to a vote.

Man: Done.

Man: So one thing that is important to me is the trend of the measure. And I see a lot of inconsistent or inadequate presentation of those data.

And I guess the question relates to what we determine to be top down and that's a slang term that we've used forever. I'm not sure it's very clearly defined.

But some of these measures that I've looked at appear to not have changed or -- in fact -- not gone down or were a three or trend, so is that an important part of this discussion? And what do we do when a page is missing about that?

Man: Actually that's a terrific question. And you're right. It's not easy to identify exactly when a measure is topped out, so we do need to rely on, you know, a lot of expertise. So we need to look at these very carefully.

And there may even be reasons (unintelligible) a measure appears to be topped out. There may be other factors that are coming into play. So, you know, process measures are different than outcome measures and patient reported outcomes measures are their own beast.

So we'll have some discussion around that undoubtedly as we get a little deeper into the - our conversation today. But it is something that's important especially when you're coming - considering whether or not to put something towards a reserve status designation or as topped out.

We do have the expectations (unintelligible) are that the (unintelligible) for your data will be presented, but it's included inside of usability, which is not a must pass criteria.

Man: Yes.

Man: In one of our prep calls we were talking about certain, you know, what does that mean when performance hasn't changed, you know, it could be that it's really hard and that it's going to take some time to change. It could be that it's topped out as you say.

So some of it is philosophical and some of it may be mythologic. But I think that's certainly up for discussion measure by measure.

Man: I mean I think it's - I'm looking at (Sherry). If there is - I'm - I'll just close on my memory (CP2), but not here in that capacity. But if a measure were - if we're concerned about measures that matter and burden then I think that's a really important issue for us to not bury in some…

Man: Yes.

Man: …sub-section. I think it's a - should be top of mind.

Woman: Yes. I'll just echo that. I mean I agree.

I think that I understand why usability is not a must pass, but it also is quite bothersome to me -- particularly for maintenance measures -- where we have information back and some concerns related to the burden associated with administering them.

Woman: Sorry. I just wanted to add something.

We use (unintelligible) not must pass up until last year. And so we've been having this debate internally about feasibility. So because it's not must pass now does not mean that in the very near future would it be this, you know, contacts and discussion is very helpful in informing our decision.

Man: So the other place where - it actually appears in two places inside the (unintelligible) our evaluation (unintelligible). So one spot is in usability, which is not must pass. But in evidence it is. So for performance gap that is considered a must pass.

So the differences between those two are a little nuance. But in usability it'll be your over your data.

Man: Right.

Man: So we're looking at how it's performance over time inside of what's called performance gap instead of evidence. And that is a single snapshot that we're - we're just looking at how providers staff against each other, the distribution, are they tightly grouped and so it's really discerning whether or not we can distinguish between the quality of two individual providers (unintelligible) and reliability.

So we're seeing this data in a couple of spots and some places it's must pass and others it's not. So we do have to be careful when we're getting to those spots of those discussions that we do indeed know what we're considering with this data.

Woman: I don't want to prolong this discussion, but I also don't want to see us getting into a very rigid position on this point. And I raise it in part, because in the world of pediatrics we have a lot of measures in which we have some pretty good high consistent performance. And over time the numbers don't change very much.

But you might want to look under the hood and find that -- wow -- the overall number looks pretty good. If you look at certain segments of the population it's the performance that does have a gap. So you'll get into these discussions. It's always going to be a little squishy.

Woman: All right. (Unintelligible).

Woman: Just to add one point. I mean I think one of the issues that I struggle with on the usability aspect is on the choice of how they're used by payers and others. I think in many cases many of these are very good measures that can be used appropriately for quality improvement and so therefore having an NQF endorsement can be quite helpful for them to understand the components that went into that.

But then on the many cases they also feel in just them being used for payment for accountability purposes can be problematic. And that's where I run into real issues under the usability aspect quite frankly. Not only related to burden, but also whether it's truly appropriate to tie payment to certain things when really that puts another level of - you know, pressure's not the right word, but

it just put it in a different category in the clinician's mind when they're seeing their patients and what they can and can't get value out of from using the measure. So it just puts it in another...

Man: (Unintelligible).

Woman: Yes.

Man: (Unintelligible).

Man: It was - maybe that can serve as an NQF historian. And I don't think anyone in this room would - they usability if it was used for quality improvement. It's what (Sherry) said. When it's get translated into this higher order of expectation that we fall down with our concerns. And so that's where I think that the friction is here.

Woman: Right.

Man: So…

Woman: Yes. I look at the use…

Woman: I will also add that this is attention and (Don) and many who have…

Woman: Yes.

Woman: …been around our table -- (Sherry) and (Chris) -- that you know we've been talking about this for quite some time. It's come up most recently with the consensus standards approval, because it is a (CSAC) about whether our

measures should be through the poll for both quality improvement and accountability.

And, you know, we've had technical expert panels talk about this or pine about it, but, you know, it's becoming increasingly tense as the stakes have gotten higher for everyone in the healthcare system.

And we realize that. We've always said that we want committees to look at these measures on their merits and the scientific merits of how they're constructed and to be agnostic to use. And I think it's almost impossible to do that in this day and age.

I think what we are trying to say is the agnostic to the (unintelligible) program use. I think it is within your purview to be thinking about accountability particularly if those measures are intended for those purposes as well, which we require.

(Sam Stopey): Okay. Thank you. It's a little after 8:45. We need to get started with the measures. (Chris), did you have one more burning thing or are you just putting (unintelligible)…

(Chris Stilla): Just a quick clarification on what we mean by topped out. I'm thinking topped out is high performance on a measure, say like a 90 or so. With a (unintelligible) with what Lee said it could be underlined things going on there (unintelligible).

But we don't mean topped out as, like, a 60, 60, 60, you know, no movements there (unintelligible) good.

Man: (Unintelligible).

(Chris Stilla): Okay.

(Sam Stopey): Okay. Thank you. (Anna), if you could put your sign down just so I know that you don't (unintelligible) want to speak.

Woman: Yes. (Unintelligible).

(Sam Stopey): Great. Okay.

We are going to start as my computer -- just decided to black out on me -- with measures 3227. The collaborate your decision-making score. And many of you may remember -- was a year or two ago -- the developers came to discuss its development a little bit with us and ask for some guidance. So thanks for that. It was a good discussion. And now they're back for prime time.

Are they here or on the…

Woman: They are.

(Sam Stopey): …phone?

Woman: The phone.

(Sam Stopey): They are on the phone from (Dartmisk).

So we'll have them introduce the measure for two or three minutes. We have an hour allotted for discussion, for the total discussion. And then the

discussants will be (Steven Lead) and then (Lisa), (Peter) and (Deb). So ready to go.

Is the developer on the line?

(Glen Alvin): Okay. Good morning. My name is (Glen Alvin) and thanks for invited us here.

Woman: (Glen), we can't…

(Glen Alvin): I also have my…

Woman: …hear you.

(Glen Alvin): …colleague…

Woman: Can you speak up?

(Glen Alvin): Hello?

(Sam Stopey): Hello. You're very soft.

Woman: Yes.

(Glen Alvin): I'm very soft? Is that better.

Woman: Yes. Much.

(Sam Stopey): Much. Thank you.

(Glen Alvin): Okay. Sorry. Well thank you very much for the opportunity to discuss collaborate.

And on the line also is my colleague, (Rachel Fortina), who has worked with me for many years on this measure.

I - I'm going to give an overview of the measure if that's okay.

The challenge that we were facing -- I think -- in this field of shared decision making is to measure something, which patients may not have experienced in any great debts in the past. So that's one level of challenge.

The other challenge -- I think -- is that decisions are not very obvious to patients obvious - often. They may not be explicate and maybe -- although every encounter I think has a decision of sorts in it -- it may have been not made obvious to the patient if there was a decision being made as do nothing or to carry on or even to compare different options for management.

The third challenge we were facing as we were making a measure here -- a patient reported measure -- was the terms such as decisions and options and preferences, a kind of language that we use in the field of shared decision making are really difficult for patients to understand.

So the first business we had really was to develop a construct of theoretically based measure, which has very easy language for patients to understand.

In terms of the theoretical construct we went back to the mobiles of shared decision making, one, which I've been working on is called collaborative deliberation. We've written about that in theoretical papers and you can find

those if there's any interest. There's also work by Gregory Makoul called, "An Integrative Model of Shared Decision Making".

And if I boil that all down shared decision making really is constructed on the bases three pieces of work in any conversation.

One is in summation provision to patients. You're letting them know about what we're doing, what the possible management options by the treatments or testing. And that can take some time -- of course -- and a very important that people have accessible information and evidence about what's possible.

And the second thing is to elicit a patient's preference. What do they make of this? How do they react to this? Do they have any priorities themselves? How does it fit into their bio-cycle social world as you were? That we call preference elicitation.

But then there's a third part, which is even more difficult -- I think -- when we look at the encounters and look at transcripts of encounters, which is to integrate those preferences into a decision made. And that I think is a (C-cornered) stool on which we've built collaborate, information exchange or provision, preference elicitation and then preference integration.

And if I just read out the items to you I think you may kind of hear those in the items. The first item is how much effort is made to help you understand your health issues. And then the anchor goes from no effort is made to every effort is made on a ten point scale, zero to nine.

Second item how much effort was made to listen to the things that matter most to you about your health issues. And that's the preference elicitation phase. Again, same scale.

And the third item, how much effort was made to include what matters most to you in choosing what to do next. And that's the preference integration step.

And -- as you can see -- we have here in a ten point scale, which is anchored the two extreme points. We also have a five point scale anchored at every point. And the second (unintelligible) between the results here was very minimal in the papers that we've submitted.

We've got some (unintelligible) and if I just try and summarize what we were trying to do with this fast and frugal measure -- which fits into another suite or tool like this -- is to be sure, is to be focused, is to be technology enabled to be used as only a smart phone or online. It takes 30 seconds on average for patients to complete this.

And really in my view trying to address these three constructs is what (Michael Barry) answers (unintelligible) trying to address the clinical of patients and of care, which is to involve patients well and ethically in decisions.

So I'm happy to respond to questions, but that's what we've been trying to do with this measure.

(Sam Stopey): Great. Thank you very much for your (unintelligible) and good presentation.

(Steven), you want to kick it off with some discussion?

(Steven Hoy): Sure. I didn't expect to go first. But I am - my initial reactions to this measure were very positive. I liked the simplicity and overarching applicability of the measure. It does cross it's references crossing different clinical interactions.

Easily my - something that spoke to me about the measure is it speaks to shared decision making as a relationship. The shared decision making is mistaken for the use of a shared decision making tool where in reality if they give and take between a provider and patient and family members, so I believe that that's a strength of the measure.

It's, you know, I think I'm just giving a brief overview of it, so - and we dive into the (unintelligible) I think I'll stop there for now. It is an outcome (PROPM) measure, which is an interest classification for shared decision making, but it…

Woman: Interesting (unintelligible).

(Steven Hoy): Yes. So I'll stop there and invite my co-chairs also to - anything else has stuck out to them.

(Sam Stopey): Great. (Lisa), you're next.

(Lisa Maurice): I really loved it. This is included. It is something that clinicians are being measured on. If they are offering and stating any shared decision making, so I think that that is potential use down the road in terms of determining quality (unintelligible) what seems provided.

I would disagree with one of the comments from the scientific panel, which was that shared decision making doesn't need to happen in every encounter. What my thought was - and the specific example they gave was you don't have shared decision making with a broken arm. And I thought it could be as simple as what color of cash am I going to get. That could be very important to somebody and that could be a shared decision. It's not - it's - the course

that it's taken clinically is not all going to be in line with what a patient is wanting. And to a (unintelligible) item is so important to know what the patient is thinking.

(Steven Hoy): Additionally if I could, I - from my experience clinical staff has a really hard time analyzing their own ability to do shared decision making. It's kind of like if you ask a hospital, "Are you patient and family centered?" Everyone says, "Yes.", right? But - and clinicians have a hard time understanding how their patients and families perceive the opportunity to participate in decision making. That's another strength in this measure -- I think -- is it actually starts to set a little bit of an expectation that patients and families will have, involvement in the measure.

I see (Sherry)'s up next.

(Sam Stopey): We'll have discussions a little bit after. I just wanted to give (Peter) and (Deb) a chance to give their general overview of what (unintelligible) (Sherry) (unintelligible).

Man: Can she go first if that's all right?

(Sam Stopey): Sure.

Man: Who's the other speaker?

(Peter Thomas): So I was struck by how direct an impact on provider behavior in performance this measure to - could (unintelligible) could create. As opposed to a measure such as a change in functional status limitation in a (unintelligible) care environment, which has - in which a multifactorial and has a lot to do with the

own - the patient's own ability to respond to a (unintelligible) program or something of that nature.

And it struck me that if providers know that they're being asked these questions after their encounter they could pretty easily change the way that they interact with patients and improve their scores. And so I thought that - I was struck by how indirect an impact they could have in quality improvement.

And I wanted to ask whether in the tests -- if I could ask the developer -- in the evidence section did the providers know that these questions were being asked or do they not know?

(Glen Alvin): Shall I respond?

(Sam Stopey): Yes. Please.

(Glen Alvin): Yes. Yes. They were aware in the first set of initials in the New Hampshire and in California and in (Kelsey) Boston they were aware.

And - but we didn't go so far as to provide regular feedback. In fact we did do that in Dartmouth and we saw some improvement and we gave feedback. But that's a step that we really want to amplify and test some more. We're doing that actually in the veteran's administration in Massachusetts and in other places and, like, go on about that.

But you're right. They were aware.

Man: I just work here in part for the consumer perspective and I happen to have some very recent consumer experience with this yesterday. I went to my

doctor and I was rushed through this appointment, I could barely get his attention and it was really (unintelligible).

If I had responded to this survey after my appointment he would not have scored well. So I think this is a very important measure.

Man: In other words there's a gap.

Man: There is a gap.

Man: Okay.

Man: (Unintelligible).

Man: I was honestly surprised at the sort testing I wrote. It seems high to me.

Man: (Unintelligible).

Man: (Unintelligible) or whatever…

Man: Yes.

Man: …to. Yes. Okay. Yes.

(Deb Soliva): I certainly agree that there is a significant gap in shared decision making and that we definitely need measures in this area.

You know, I go to the - have seen - go to the Los Angeles Philharmonic and when they get up there to perform I don't think there's a lot of effort.

So the word effort does sort of bother me a little bit in the language. But -- that said -- there's a huge gap. This taps on the three elements of shared decision making. I just think that some providers come across as, you know, very interactive and very calm and very, you know, and I give effort to may not come across.

But -- again -- I think, you know, we need measures in this area. So I think there is a gap.

(Sam Stopey): Great. Okay. Thanks. So go ahead, (Sherry).

Woman: Thanks. And I do fully agree there's a gap here and it's something that we have really pushed forward for our members to engage more into decision making and trying to figure out the best way to offer them tools and information and education around how to do that.

And I guess that's what my question is related to this. You mentioned that (unintelligible) for the developer the feedback. Some feedback has been provided and there's more testing going on along those lines. I would be concerned about this measure being implemented for accountability purposes without actionable feedback being provided in such a way that the clinicians would be able to take what they're hearing and actually make - be able to make meaningful changes in what they're asking.

So there's a difference between, you know, honestly asking somebody what color cash they want versus truly engaging in a shared decision making discussion that does involve training, that is not received in medical school.

(Sam Stopey): One or two more general questions and then we'll start to go category by category.

Woman: So I would just point out that in any quality improvement effort there should be some kind of feedback loop and measurement is just one aspect. How people take this and we see this in the other measures where we may not have seen a change -- for example -- in scores over a number of years, but that's more a function of people not taking the information and using it for feedback. That's not part of the measure itself though.

Woman: It is some kind of feedback (unintelligible).

Woman: Well - but the individual practice needs to have access to the data, so that they can then take that data and feedback and implement change if appropriate.

(Sam Stopey): Sure. So that's usability and we'll talk about that a little more in a minute.

(Don), did you have one more general question?

(Don Casey): Yes. Thank you very much. In fact I was just up in the (Handover) last weekend and it was nice.

I - my question is that - I - I'm not familiar with this, so that I remember - I think you've taken to NQF about this a few years ago.

This is measuring a shared decision making process in your own words, which I think is important per what (Sherry) said. But in my particular case relative to hyper tension what we're working on specifically is shared decision making around a clinical decision as to how you want to reduce your risk of cardiovascular events.

So I'm much more in the weeds here and this doesn't seem to have the follow on question of what happens next or what did happen relative to the impact.

One could argue that shared decision making is used perhaps inappropriately for low value care. In other words I'm going to do this, I mean I want to get some experimental treatment, because (Joe) had it and (Joe) feels better for two weeks.

So I would just be careful about this. I think that the trend is good, but from the standpoint of moving this into a public reporting and payment standpoint I think it's just measuring of process, not an outcome.

(Sam Stopey): Okay. All right. (Ellen Sholts), one last thing and then we'll get into the…

(Ellen Sholts): Just very briefly. I want to push back on a little bit. A real (unintelligible) shared decision making (unintelligible) information that this is experimental and there's not an evidence based (unintelligible) decision around that.

However, I also noted for the (unintelligible) like, a process measure rather than an outcome. So we don't necessarily have to get into this, but I am (unintelligible) from NQF if you have a clear definition of (Pro-PM)?

This is not the first time I've seen examples on measures that kind of want that (Pro-PM) label. It's the exciting new thing the direction measurement is going in. Sometimes I think it gets a little broader perhaps than it should be.

(Sam Stopey): Yes. Go ahead, (Susan).

(Susan Sparcet): We do have patient reported process measures. They are rare. I'm not sure that we have any in the portfolio. But you could have a patient reported process measure if that's what you're getting at.

Man: So I - we agree with you that this shared decision making is a process, a patients perception of the shared decision making. Now we're getting a little meta. But that is why this would qualify as a patient reported outcome. It's their perception of care and a sense of involvement and that would be considered experience of care/patient reported outcome of perception of care.

And we do have standardized definitions around these. The - some of them are dated back as far as 2012 where we issued our initial report around (Pro)s, (Prom)s and (Pro-PM)s, but it's also been updated inside of our current work as well.

Man: And if it's done right -- I mean -- this could have a material impact on outcomes, on - you know, I mean this is a real - this could really may be an (unintelligible) play on whether patients get - and go in one direction or another in their treatment plan. I mean it could have significant outcome implications.

(Sam Stopey): Okay. Good. Thank you for this. And (unintelligible)…

Woman: Yes.

(Sam Stopey): …one more?

Woman: Just a question. I noticed in the write-up that it is currently being used for accountability through BlueShield of California and I didn't see any discussion of - if specific payments were attached to this or not. And so I'm

wondering how deep it is into an accountability model and how that's working as opposed to feedback to the provider, et cetera.

(Sam Stopey): Yes. Could the developer respond real quick to that, what you know?

(Glen Alvin): Yes. It's not part of a payment model. We know that for sure. It's being used as a patient reported measure in terms of their preparation for Orthopedic surgery.

They've required -- before people accept and (unintelligible) of big surgery for hip or knee -- for people to score well on the collaborate score.

And but - yes. And they haven't published on this data and - although they have made conference presentations. So we haven't gotten much visibility into how they've used it for accountability. But that's - the settling of payments involved there, but they've been doing it as a performance of the Orthopedic surgeons.

Woman: Noting that in the report it says that BlueShield of California is using it in a payment program. So you - I'm just making that note that that's…

Man: Yes.

Woman: …what we - the information we were given.

(Glen Alvin): Right. Yes. And I have no more information on that myself. They haven't -- as far as I'm aware, (Rachel), -- published on the…

(Sam Stopey): Okay. (Don), did you have something else here?

(Don Casey): (Unintelligible).

(Sam Stopey): Okay. Great. So let's start to talk about the evidence and things like that.

There were not a lot of concerns that (Steven) and then the other discussions, anything?

Man: I'm hearing the committee is kind of overwhelming discussion about the - that this may pass a evidence in opportunity for improvement. But (unintelligible) welcome for other thoughts.

Man: I probably talked about evidence. Yes. I mean the question - the questions have been covered really that…

Man: Yes.

Man: …from NQF staff to us. (Unintelligible) one thing the provider can do to achieve a change in the measure is always a good discussion of that.

Does the target population value the measured outcome and find it meaningful and I think we've had two pretty a discussion about that. So…

(Sam Stopey): Anything else (unintelligible) (Don)?

(Don Casey): Well - sorry. I'm looking at 183 and 184 and they're blank. So I understand that - I get it confused, because I may wonder why they're there in the first place, if they're supposed to be blank.

Man: So for 183 there are certain portion of evidence submission where the developers only asked to answer one of the questions, options depending on

the measure type and the type of evidence that they'll be submitting. So it's actually appropriate to have some of those blank.

(Sam Stopey): So do we need to go to a vote then on evidence and then discuss a little bit more? People may need a minute or so to get those links active. I know I do.

Man: (Unintelligible) some of the emails - I'm sorry.

Woman: It's in your calendar invite.

Man: Oh in the calendar invite.

Woman: So if you open up the…

Man: Okay.

Woman: …calendar invite you should see the link for voting.

Man: Okay.

(Sam Stopey): Yes. (Terry), go ahead.

Woman: (Unintelligible).

Man: Hi. (Unintelligible) I'm going to refuse myself from this measure. I'm based at (Mass General), which is one of the sites that this is involved in. It's creation (unintelligible).

(Sam Stopey): Thank you.

Woman: (Unintelligible).

(Sam Stopey): And (Sherry) and (Linda) are recusing as well, so voting is now open on NQF 3227 Collaborate Shared Decision Making Score, (1-A) evidence (unintelligible). All right. Now (unintelligible) can you help (Brian) (unintelligible). He needs a little bit of help.

Woman: So we're looking for 18 votes. And it looks like we have 16. One other person's vote has not registered unless there's someone who stepped away.

Man: (Unintelligible).

(Sam Stopey): Right. (Unintelligible).

Man: (Unintelligible).

(Susan Sparcet): (Jordan), can you pull up the slide with the wireless information. We do have a wireless network if anyone's having trouble connecting to that. Thanks for your patience everyone.

We should have - okay. Has everyone - are you sure you all voted? We seem to be having one vote not registering. Can (unintelligible). All right. So the measure path is evidence 16 yes, one no.

(Sam Stopey): Okay. Now we move onto performance gap. We just have a little bit of discussion about that. But I heard a hint or two that people might just want to have a little bit more.

Any concerns about -- yes -- all the questions that I've heard have been pretty much positive. (Deb).

(Deb Soliva): I think the initial data that we (unintelligible) a couple of years ago would indicate that there - it was almost topped out. But then there's new data in here that shows that there is a performance gap (unintelligible) data in a larger sample. So there is -- based on the evidence that's in front of us -- there's a performance gap.

Man: Yes. That - the range that I saw was .68 to .8671 with the same deviation of .09, so reasonable.

Man: And it says note some disparity in the data. That would be very important for quality improvement efforts.

(Sam Stopey): Okay. Any other discussion on gap before we vote on gap?

Let's vote on gap.

Man: The voting for NQF 3227 for performance gap is now open.

Man: Okay. We got 17 last time, we have 17 - now we have 18. Good. With 18 votes the voting is closed and performance gap passes.

(Sam Stopey): Okay. Now we have the scientific acceptability. The scientific methods panel essentially passed unanimously. (Sherry), anything that you remember from - anything you'd like to say about that?

Woman: Actually since I'm a consultant I might - yes. I have good…

(Sam Stopey): Okay. Good enough.

Woman: …(unintelligible).

(Sam Stopey): All right. Thanks. Okay.

Woman: (Unintelligible).

(Sam Stopey): Do we have anyone else who's on the scientific methods panel in our group?

Man: I don't think so.

(Sam Stopey): Okay. All right. So…

Man: There were some notes about the low response rate in survey minimum. I don't know if anyone bothered to talk about that recently.

(Sam Stopey): Okay. Right. It shows some of this was implement ability and usability that I think we've talked about is going to do - is - this is unclear as to whether everyone who should have been eligible was surveyed. That was really the only concerns that I saw.

Go ahead.

Woman: I think the question was how the minimum number of (unintelligible) individuals need to be selected. And the providers gain that by selecting for the people they spent a lot of time with in discussions (unintelligible).

Man: (Unintelligible).

(Sam Stopey): Sorry. I (unintelligible) to that. Any word from the developers on how that went during the testing?

(Glen Alvin): Yes. We were very careful in the initial testing ourselves. Not to allow the providers to select or give a survey to the patients. It was sent to the patient or given to the patient by a third party outside of the encounter room.

I think that is a very important concern when it's used in the wild and I think we would always advocate. Either a technological invitation or administration or that we would not allow the clinician themselves to field the survey to the patient in the consulting room.

If - even the scores -- when it was done within the practice -- were higher actually. And so when we achieved (unintelligible) of data collection we achieve lower scores.

(Sam Stopey): Good. And I think that's consistent with a lot of other measures like that at - in other places.

(Sherry)?

(Sherry): This is a general question. In general the patient reported measures like (CAP)s and other measures are intended to be often fielded by a vendor to be - to get out of the (unintelligible) the system the same way and so on.

So this is sort of a general question. Is the assumption -- unless dated otherwise -- that this will be some kind of external body that ultimately implements these kinds of measures or is it measure by - are we to consider these measure by measure? And does the measure developer then have to clearly state their intent on how these will be administered?

(Glen Alvin): Yes. Can I respond to that?

We can make this (unintelligible) open access under creative common license. So people could be free to use it of themselves under quality improvement. But we are actually working with a vendor such as those -- and I don't want to mention them here -- but would field it as a third party.

And that -- I think -- would be the - it would need to be stipulated at the (unintelligible) uses of an accountability measure. You would have to avoid -- I think -- the bias of provider administration.

(Sam Stopey): Any comments from the NQF staff about (unintelligible)?

Man: (Unintelligible) weigh in.

So the - measure (unintelligible) are free to add recommendations of this nature inside of their measure specifications. However, if it's really an implementation issue where since you are relying on the expertise of those who are going to be using the measure in -- as the developer said it -- in the wild. I really like that term.

And so it's not necessarily something for the committee to take under too much consideration.

(Sam Stopey): Okay. Then just a reminder to (unintelligible) we're still talking about the scientific examples. Thanks (unintelligible). Other comments can go in the next thing.

But, (Don)?

(Don Casey): Yes. I'm reading your article from 2014 Dartmouth's and it says that to detect an estimated 15% difference in the top score you need 216 participants. One reference who sided was 30 participants. So could you comment about the (N) that's needed to determine significant differences? Fifteen percent -- I think -- I assume would be a significant difference, but can you comment on that?

(Glen Alvin): Right. And I think we're maybe talking about two issues here. And we don't feel comfortable estimating the nation level score until we've got 25 responses tucked in the shin of the (unintelligible).

But I think if you were to scientifically say that there are differences between groups we would go then to the kind of (unintelligible) statistics and say we need that power to make a defense at the group level. I think slightly different issues there.

(Rachel), do you want to comment further on this?

(Rachel Fortina): Sure. I just think just to add in our reliability (unintelligible) discussing we did see a sample probably around 200 having a minimum - meeting a minimum of liability standards.

(Sam Stopey): Great. Thank you. (Deb).

(Deb Soliva): I had a couple of questions, but I just needed help with.

(Unintelligible) (ICC) was .012? That was reported in here. But no one on the message panel really commented about that.

And the second was that on the sensitivity analysis that the performance score reflected shared decision making and 39% of clinical (unintelligible) where all

of the elements were present in the vignette. And I'm not a huge fan of a vignette (unintelligible) a way to do validation. But I just wondered if there were - was any kind of - if the developer wanted to comment about why they thought that was so low.

So the first is help me with the (ICC) somebody.

Man: (Unintelligible).

Woman: Okay. (Unintelligible) this - the (unintelligible) relation coefficient tells you how much between versus the - if the ratio between over - between plus within unit variation areas.

And so the within unit variation is error. So if there's a strong -- for example -- if there was a physician thumb print across all patients and their practice and there's a lot of difference between physicians, so the (unintelligible) down here, a little standard air (unintelligible) are tight, so you can tell the difference between one group and another.

If there's a lot of within physician variation that says there's not a strong physician thumb print and therefore at the physician level you worry that given the function of what the doctor's doing it may be a patient level variable.

So I - it's a legitimate -- in my view -- that's a reasonable question to be asking if it's to be used at -- for example in this case -- the physician.

(Sam Stopey): Thank you, (Sherry). I was thinking we - you should do a video that we all watch about psychometrics. (Unintelligible).

Woman: But, (Sherry)…

Woman: Yes.

(Deb Soliva): …I actually - (Sherry)…

Man: No. No.

(Deb Soliva): …I asked that question, because I do understand what the difference is…

(Sam Stopey): I think it's great.

(Deb Soliva): …supposed to be representing. (Unintelligible) I'm trying to understand where a .012 sort of falls within that.

Woman: Well that basically tells you there is not much - there is not as much between. That ratio is small.

(Deb Soliva): Right. That's…

Woman: There is not as much…

(Deb Soliva): …what I thought.

Woman: ...between versus within…

(Deb Soliva): Okay.

Woman: …physician variation and if you're trying to achieve a level of comfort around how much the liability there is I would ask you back would that be enough --

without getting into the weeds on this particular measure -- would that be enough to give you…

(Deb Soliva): I (unintelligible).

Woman: …confidence that -- yes -- in fact it's reliable at the level it's being tested.

(Deb Soliva): Yes. And I think part of the reason I was asking is it just seemed lowish to me, but nobody on the message panel seemed to bring it up. So that's why I was, you know, a little confused.

Woman: Well that's - without -- again -- getting specific on this (unintelligible) that's one of my issues about some of these testing is that when it's actually measured at the units being - and it comes up a bunch in the things that I was looking at. That makes me uncomfortable. I don't - guess that's the level of the liability .01. Ouch. That's a - that is a lot more within than between unit variation.

So -- yes -- that makes me uncomfortable and my colleagues at the - at NQF have heard me go off on this issue before, but NQF was at the time a measure was being submitted I think had a different guidance about how those were to be interpreted and some of those (unintelligible) issues came back on what's half of the liability, which is not what's intended to be tested.

So…

Woman: Yes.

Woman: …I think that the…

Woman: And, you know, it's also possible you and I had this argument before.

Woman: Yes.

Woman: That, you know, you could have a provider who treats patients differentially.

Woman: (Unintelligible).

Woman: You know, say is treating patients of certain races better than others.

Woman: Right. Right.

Woman: And so there could be a variation at the provider level that's actually legitimate.

Woman: But I just thought this was low and I surprised…

Woman: You've used (unintelligible)…

Woman: …that…

Woman: …(unintelligible).

Woman: I mean I would view that as low and I didn't mean to impute - encourage your…

Woman: Yes. No. No.

Woman: …interpretation of (ICC). But -- again -- if there is more within physician variation then it's used to discriminate between those issues if that's the goal.

Woman: Yes.

Woman: Okay. Then that's going to be compromised by that level of what would be considered noise.

Woman: Thank you.

(Sam Stopey): Great. And was there another part to your question that you needed…

Woman: No. There's - it was actually a second question. I asked two questions combined. The other was just the (unintelligible) analysis, 39% of the clinical vignette. If there was a correlation, the 39%.

(Sam Stopey): So do you need the developers to…

Woman: Yes. I wanted the developers to just sort of comment on…

(Sam Stopey): Okay.

Woman: …what they thought may be going on, you know, why there was a low percentage of (unintelligible).

(Glen Alvin): So -- to get back into my memory here -- this was published -- I think -- two years ago.

We - the - and I agree with you that the stimulated study is not ideal for psychometric testing. But I think it was an important first step for us to demonstrate that.

When we've got a (unintelligible) constructed vignettes to have very low levels of shared decision making, medium levels and very high of (unintelligible) control that exactly and I think if you - I don't know if you've had a chance to look at the graphs on the paper, we do actually see a very clear response in the internet panel that we used and so collaborate scores (unintelligible) sequentially as we improve the level of shared decision making in the audio tape simulations. But from that point of view I think we were very happy with that.

There was a bit of noise when -- I think -- we got to all that the - there was lack of precision when we got to medium and high levels. The measure couldn't distinguish very well. But between the low and the high levels it was discriminative ability.

I'm not sure if I'm answering your questions directly though. And, (Rachel), do you want to comment…

(Rachel Fortina): Yes.

(Glen Alvin): …some more here.

(Rachel Fortina): I think that worries me just a tidge. The idea that, you know, it - a lot of measures do a good job of discriminating (unintelligible) extremes. But I'm not sure it's going to be limited to use of the extremes.

And so if you're telling me -- which I didn't really clearly see in the data -- the ability to discriminate between moderate and (unintelligible) practice is limited, then that's a little troubling.

And I still - I mean what it says here is that 39% of the vignette - let's see. It says, "The collaborate performance score reflected (SDM) and 39% of the clinical vignettes were all three dimensions of (SDM) were present. So this is like an idealized scenario and - where you've tried to put all three elements there and only 39% of the time did the (SDM) agree, which seemed a little - but maybe I just don't understand what happens in the study.

(Glen Alvin): Yes. Just a comment on that is that in (unintelligible) studies when we look at recordings of actual practice shared decision making levels are at the floor usually and not at the moderate or high level. And so I think when we're - we take this measure into wide of practice, and when I think there's a chance of that when we get wider use after endorsement, we will probably see that an actual practice levels of shared decision making seem pretty low. And then we will have a better sense of the clinician level and of the group level of where we're - where we have the low sensitivity to the measure.

(Sam Stopey): Thanks.

Woman: Thank you. And that's true. I mean I would agree that - I think we all agree there's a gap. And I will also say in your data you're highly correlated with the (CAP)s measure, so that's something that is reassuring…

(Sam Stopey): (Unintelligible).

Woman: …on - in your data.

(Sam Stopey): Okay. Great. (Chris), to show a point and then we have to move.

Man: This is scientific. I'll (unintelligible) space to my understanding or belief that shared decision making really needs to rely on foundation of (unintelligible) literacy and patient engagement in my opinion.

I'll (unintelligible) reliability, but if the patient had - the folks who filled out the survey -- which expresses what their feelings were about effort -- did they have any understanding of what shared decision making is? And that - you know, it goes back to a comment I heard earlier about decisions aren't necessarily made every visit, but frankly the - these visits and the interchange with patients and the engagement really helps with decisions down the road too, which is shared.

I guess was there - is there any way to know what -- if you're asking - if I'm filling out the survey and I express what I feel about shared decision making -- do I really know what that is.

Woman: (Unintelligible).

Woman: (Unintelligible) the questions do not specifically say shared decision making. The questions are asking about the encounter, but not using that language per say. So I'm not sure that's as important in - but I really appreciate your concern around what patients do and do not understand.

And if they understand those questions then that, you know, that goes to the reliability.

Man: I guess my point is I wonder if this is more of representation of - in the patient rather than the provider. And I think that goes to the…

Woman: I…

Man: …discussion we were having about the (ICC) so far. Okay.

Woman: (Unintelligible).

Man: Just…

Woman: Yes.

Man: …one more quick thing and then…

Woman: Yes. Because I messed up the response to (Debra).

In patient reported outcomes that are multi-measure when they're combined then these additional error term on the denominator is within patient across a item.

So you've got that source of error. You've got two sources of error in the denominator. So the (ICC)s for multi-item measures are going to be lower and - than you would expect to see, but that low is a problem.

(Sam Stopey): Thanks. (Ellen Sholts), did you have another new point?

(Ellen Sholts): I just very briefly want to share a perspective. I think (it fits) under reliability and specifications is that I have to say this is probably one of the most holistic measures I have seen, like, particularly come to NQF and that had applied to any kind of close encounter in any patient and we need more of that.

When we all talk about how many different measures there are -- I mean -- think about how much time we all had to spend reviewing measures. It's a lot

of measures, right? And we (unintelligible) fewer more meaningful measures to me, this reflects that.

Man: Well - and then I would say -- if anything more holistic the measures a bunch of different stuff at the same time would have slightly incrementally (unintelligible) liability as well.

Okay. I think we need to vote on reliability and validity to try and keep some somewhat to time.

Yes. So the vote is to accept the scientific methods panel, which was moderate I believe, right? We have 18.

Woman: Good.

Woman: Okay. We just have to have consensus that we're accepting a methods panel vote, not (unintelligible). So -- yes -- we'll take the methods panel recommendation and reliability. Fourteen yes, four no.

(Sam Stopey): Okay.

Man: Yes.

(Sam Stopey): Great. And then I believe we have the same question for validity come up in a minute. Do we need to do the two eight (unintelligible)?

Man: No.

(Sam Stopey): Maybe just (unintelligible) that. Got you. Okay. (Unintelligible) votes for - great. So do you accept the scientific methods panels rating for validity?

Woman: And what was that rating?

(Sam Stopey): Which was - I believe also moderate, right?

Woman: (Unintelligible).

(Sam Stopey): Okay. So the consensus…

Woman: It is moderate.

(Sam Stopey): Okay.

Woman: (Unintelligible).

(Sam Stopey): We need four more votes. Two more votes.

Woman: Two more votes. Can everyone just make sure that you're vote has been clicked.

(Sam Stopey): Almost.

Woman: Okay. I guess we'll close it at 17. That's 13 in favor - oh. We had one and then we lost it.

Man: (Unintelligible).

Woman: Well we have 13 in favor of recommending the methods panel vote of moderate, four no. So we will go with moderate.

(Sam Stopey): Okay. Great. Feasibility is the next measures available in electronic (unintelligible). We've had some discussion about which might be fast and how the administration might be fast. Discussion from the discussions (unintelligible) others?

Man: Yes. This one is available electronically, reported that it can go - be completed in under 30 seconds (by patients). I didn't see a time on the pre-evaluation reports from the committee, but the question being is the burden of patients and providers to administer the survey outweigh the value of information gained. No. No. No.

This is important to (unintelligible) populations to on (unintelligible) adversely it would be easy to implement and, you know, it's quick and easy, so…

(Sam Stopey): Great. We have four. (Don) was first and then we'll just do (Don), (Beth), (Theresa) and then (Sherry).

Man: Just quickly, remind me what the minimum sample size was for the provider.

(Sam Stopey): (Unintelligible).

Man: …did I hear 25?

Man: Yes. I believe it was 25.

(Sam Stopey): Yes.

Man: And we don't know the interventions past what the data shows, so - at least I didn't see any.

Woman: So one question as far as future use is there any (unintelligible) with another survey or be incorporated into (CAP)s -- for example -- as far as feasibility, otherwise (unintelligible) we might be competing for patient attention (unintelligible) two surveys and also the cost of a vendor for two separate surveys?

Man: Yes.

Woman: So just a question…

Man: Yes.

Woman: …I think these are great questions (unintelligible) be able to be incorporated in…

Man: Yes.

Woman: …(unintelligible) surveys.

Man: That's a good question. I know there's some pediatric surveys out there that are not, you know, NQF endorsed measures. But the same - very similar types of questions are asked in those larger surveys.

Woman: So I just thought that - and (unintelligible) wrote this (unintelligible) the scientific methods panel rated this as moderate on feasibility, but I think the benefit is so high relative to answering three questions being not that hard that I would rate it high. Just saying.

(Sam Stopey): Yes. And I think maybe that's more usability, but that's okay.

(Sherry) and then (Ryan).

Woman: Thanks. My one reaction to that was that it only has a benefit if the physician is - has an ability to know what to do with it. And so that I think is a challenge.

But with regard to feasibility I know I'm standing from a developers, administer by a third party and maybe this fits a usability too, but in terms of the cost of those (unintelligible) unsure that it is administered appropriately and also is less burden to the clinician. (Unintelligible).

(Sam Stopey): Thank you. (Ryan).

(Ryan Collar): Thanks. I just wanted to clarify. Is the intended reporting vision for the measure (2-D) reported at the clinician level or at the clinic level?

And I ask in part related to the (ICC) discussions, because I think also within a clinic or a system obviously there can be wide variations provider to provider and how they perform on the (unintelligible).

(Sam Stopey): That's a good question. Developers have any insight about that?

(Glen Alvin): Yes. We're (applicating) here that this is used for at the clinic level. Until we get more data the clinician level.

So we - and also one of the issues that we haven't quite addressed is what percentage of the clinicians at the clinic would have to be included in order for it to be a reliable clinic measure. I think there's more work to be done there, but this is definitely clinic level at the moment.

And I think we are actually having a lot of interest in the clinician level, but we are using that as a quality improvement and using a feedback on the evaluation, the impact of feedback and not just to do more work on that issue.

(Sam Stopey): Great. Thanks. That's depth approach seems to make a lot of sense.

And did you have a question?

Woman: Yes. I was very interested at the disparity section and I don't know if this is the right time for that, so guide me, but the disparity section said people who had an interpreter scored lower. And the conclusion that was drawn is that therefor they experienced less shared decision making. And I think that's a big leap, because the use of interpreters can change how care is provided.

So I just was concerned about that conclusion that they were drawing and I don't - I hope that's not driving some of the accountability and quality.

(Sam Stopey): Sure. Yes. I think that's (unintelligible) very much an open question. There's lots of national surveys that have diverse populations and some of it is a population, some of it the language, some of it - I mean that is an open question honestly.

Great. Good. Are we ready to vote on feasibility -- I think -- because we're just about out of time.

Woman: We need two more votes. Oh, one more vote.

Man: Okay.

Man: Voting is now closed for feasibility for 3227 and feasibility passes.

Woman: Okay.

Man: Oh (unintelligible) six high and 12 moderate.

(Sam Stopey): Okay. We've had a lot of discussion on usability already. I ask that any comments be (unintelligible) have not already discussed although the discussion has been really good, so…

Woman: Okay.

Man: One thing I wanted to mention is - and that is a lot of businesses are using this kind of, you know, press one if you'd be willing to take a quick survey at the end of your 30-second call being routed through.

And I'm - I've gotten to the point where I'm being asked it so much that I just say, "No. I don't want to."

How susceptible would this measure be to that kind of dismissal by patient?

Man: Absolutely. But and the testing so far - how has that gone, developers?

(Glen Alvin): I don't (unintelligible) reached that level of saturation and I completely understand the comment by the panelist.

We recommend (unintelligible) measurement of this issue. You wouldn't want to give every patient this every encounter with the (unintelligible) care (unintelligible). So we recommend a six month (unintelligible) at

(unintelligible) assessment of the clinic level. But -- of course -- we haven't seen (unintelligible) yet.

But we definitely would not want every patient to get this at every encounter.

(Sam Stopey): Great. Will be really interesting to see how things go - as it goes. Good.

Any other usability in the use comments? Great. Let's vote.

Man: Voting for use on 3227 is now open. Your choices are, A, pass or, B, no pass.

Voting is now closed and use passes with 13 pass and five no pass.

(Sam Stopey): Okay. Let's see. What else do we need to vote on? I think just overall.

Woman: Yes.

(Sam Stopey): Right.

Woman: Usability.

(Sam Stopey): Oh usability. Sorry. Usability. Those were together. Got you.

Okay. Right. Any other discussion on usability? Okay. (Unintelligible).

Man: Usability for 3227 voting is now open. Your choices are, A, high, B, moderate (unintelligible) or D, insufficient.

Man: Really interested in how this looks changed. Okay.

Woman: (Unintelligible).

Man: Now (unintelligible).

Man: Voting is now closed and usability passes with eight high, five moderate, four low, one insufficient.

Man: (Chris), could I point of clarification…

Man: Yes. The first question was really around usability for public reporting and accountability. And the second is really about more long improvement and generating new evidences that - does that right (unintelligible) on this? I want to be sure I understand this.

Man: Yes. No.

Man: I think…

Man: (Unintelligible).

Man: …how I voted.

Man: Yes. No. That's a really important question given what we've been talking about.

Man: Right.

Man: Just want to be…

(Susan Sparcet): But the first one is how and where, if it's in use or plans for use. The second is how are the results being used to improve results and how a feedback being given to those being measured for - and how is that impacting the measure.

And then also the unintended consequences of measurement for that measure (unintelligible).

Man: Yes, (Susan). The first - when you said uses were on the first one it was for use for accountability and public reporting, right?

(Susan Sparcet): Yes.

Man: That's what you mean by use?

(Susan Sparcet): Yes. Is it being used?

Man: So…

(Susan Sparcet): Or are they planning to use it?

Man: Specifically to just plan for use. Okay. And then the unintended consequences, I'd like to get in more to your point, (Don), which is I think we see reflected in the differences in both - on one versus the other.

Man: Okay.

Man: Okay.

(Sam Stopey): Okay. Great. Overall suitability.

Man: Voting for overall suitability for endorsement for 3227 is now open. Your choices are, A, yes or, B, no.

Voting for overall suitability is now closed. The committee recommend NQF 3227 for endorsement with 14 yeses for - 14 votes for yes and four votes for no.

(Sam Stopey): Great. Good. Thank you for a great succinct discussion. We are really well disciplined, got a lot of things out there. Let's take a break and come back at 10:00.

Okay everybody. We're getting towards the end of our break. Why don't we reconvene, so finish up your conversation and we'll go ahead and get started with our next measure in just a moment.

((Crosstalk))

Woman: (Unintelligible).

Woman: (Unintelligible).

(Sam Stopey): Great. If everybody could take their seats we're going to start off with item 3461, functional status change for patients with neck impairments.

The developer is (Soto).

Man: (Unintelligible).

(Sam Stopey): And thank you for being right here on time. Thank you.

And I'll open this up here. Actually what am I going to do? I'm going to open this up here.

So we'll have a two to three minute discussion. If you could introduce the measure?

Man: Yes.

(Sam Stopey): And just push the right button, then when it turns red you're on. And then the discussions will be (Sharon Cross), who's going to lead and then (Sherry Ericson), (Ann Monroe) and (Ellen Sholts).

(Deana Hays): Okay. Thank you. Good morning. My name is (Deana Hays) from (Photo) and on the phone we have Dr. (Daniel Dwisure), who's our scientific lead on the project.

(Daniel), are you there?

(Daniel Dwisure):: Yes. I'm here. Can you hear me?

(Deana Hays): Yes. Thank you. Additionally (Daniel) and I would like to acknowledge Drs. (Karen Cook) and (Michael Callin) who are not here today, but they served as psychometric consultants on the project.

(Photo) provides a system of risk adjusted patient reported outcome measures targeting the construct of physical functional status generally for multiple orthopedic and neurologic conditions as well as Lymphedema.

Number 3461 functional status change for patients with neck impairments is a patient reported performance measure, patient reported outcome performance

measure. The two main components performing the basis for this (Pro-PM) are the neck functional status of patient reported outcome measure and a risk adjustment model for the sake of assessing risk of adjusted change at the patient level and comparing performance across providers.

A bit about those two components. The neck (Pro) measure was developed as a condition specific measure in response to feedback from patients and providers who said that they did not like these general questions being asked by a condition general measure that was being used at the time.

So in response to their feedback we gathered conceptual development, information for functional questions from a panel of physical therapist who work day to day with the patients and -- of course -- and in physical therapy the focus our standard of care is that with every patient all day long we're focusing on their function and what's important at our function.

So this panel allows us to harness the input of hundreds of patients with their input about what's important to them with respect to their function. Of course recognizing that that's not completely sufficient a follow up study included surveying patients indirectly. And patients reported the degree to which the questions were meaningful or important to them.

The items were tested and calibrated using item response theory methods. Therefore the items may be administered using computer adaptive testing for the sake of reduced burden for the patient and the provider.

The risk adjustment model accounts for eleven main categories of patient demographics and health characteristics.

Briefly the neck measure has been the focus of purely viewed publications and was implemented in 2016 into clinical setting. A recent count showed active use by roughly 13,000 clinicians in 3840 clinics across all 50 states and the District of Columbia.

(Daniel) and I would like to thank you for the opportunity to participate in this process today.

(Sam Stopey): Great. Thank you. So -- just as a point of a clarification -- this is how much of a pain in the neck do you have rather than how much someone is causing you to be a pain in the neck? Sorry. I couldn't resist it. I promise I'll stop.

Okay. Great. So, (Sharon), if you'd like to lead it off, the discussion.

(Sharon Cross): Well I - coming from the - a place where we use a lot of patient feedback I just really want to thank the developers for including that in this measure. And, (Chris), you've told my joke, so I'm not (unintelligible).

Woman: (Unintelligible).

(Sharon Cross): So that I have a whole lot (unintelligible) to say, so - but I did want to ask a question about the fact that it was mentioned that this is a competing measure with the 0428, the functional status change and she did mention that the need for a patient for having one specific connect. Do we take that into consideration at all as we look at this?

(Deana Hays): We're not looking at the other measures today.

(Sharon Cross): Right.

(Deana Hays): So we can't really do a head to head…

(Sharon Cross): Okay.

(Deana Hays): …comparison unfortunately. And we wouldn't do that anyway until both measures have been recommended by the committee.

(Sharon Cross): Okay. So there were a couple of other notes that I had just in general that I'm sure we'll discuss as we go through. One was regarding potential cost to providers. There were some conversations or some listings of information about the scientific methods panel having several concerns with this measure when it was previously submitted, but that it seems most those were resolved through the course of the review.

And I also noticed something about there was a public member comment that I thought was interesting. Are we allowed to bring that into the discussion at this point as well? We can? Okay.

So I am hoping that perhaps developers or someone else - I'm not as familiar with the terminology that was used in this but, the public member comments at the very end that we received were encouraging. And measured (unintelligible) to use the standard terminology such as LOINC is the acronym used, for coding the FYN instrument in their measure without this measure of standardization, interoperability, will be a perpetual challenge and impacts the ability to ability to measure a patient's functional status across the continuum of care.

So I didn't know if that was something that we should consider as we're having discussions about this in (unintelligible).

Man: It is my idea, probably in the usability category I'm guessing.

Woman: Okay. That's all I have for now.

Man: Okay, great. Let's see. So other - so let's see, I think (Sherry) would be next, for just sort of general comments.

(Sherry): Sure. And thanks for the opportunity. I would say, you mentioned several Peer Review publications. And I would like to hear a little bit more about the evidence. It wasn't - in many cases I found that it was referred to as preliminary or somewhere in the background.

So but you are referencing that there are public - you know, Peer Review publications related to this. And so I'd like to hear a little bit more about that. And then also a couple of other thoughts before I'll jump back over to you.

There were - I guess this gets at the - what (Sharon) brought up I think, with regard to the Scientific Method panel discussion related to I guess, probably more - I think it was more validity, but also reliability whether in some of the concerns that were raised there. Discussed because it said they did not reach consensus on their validity vote, if I recall correctly.

Man: Okay. We should probably talk about that validity part.

(Sherry): Okay, yes. So those are two questions. And then just in terms of feasibility going through these each, regarding the time and costs clinicians may encounter when using the measure. And whether or not (unintelligible) is administered by a third party and the costs that may be associated with that.

And then it's mentioned, I think, that it's not currently used in accountability programs but, plan to be. And so if there's any more that you could say about that, that would be quite helpful in terms of what its plans are and, at what level of accountability. Whether it was the individual clinician or at the clinic or practice level. Thank you.

Man: Great. (Ann), you're up next for General Comments.

(Ann): Well several of my comments have been taken, including the joke so.

Man: Wow, great minds think alike.

(Ann): But I do think, you know, I'm not a particular fan of site-specific measures. I'm also not a particular fan of body part measures. And I'm concerned, as someone who has neck pain, I also have back pain. And is this measure going to be used?

I can imagine being there at the podiatrist or the physical therapist, and they hand me a survey for neck and then they hand me a survey for back. And I just wonder how - I don't see anything wrong with the measure, but how usable it really is in a patient with multiple problems. So that's my comment on this.

Man: Great. Thanks, okay. And (Ellen), you're last but not least.

(Ellen Sholts): Well, so I definitely want to echo that (Ann). I mean I think what really stood out to me about this measure is this carving up of patients into, you know, body systems and conditions. Because it's something we do all the time in healthcare, and so it's not unique to this measure.

But I'm going to go on record as saying we should stop doing that. We wonder why we have thousands of measures. We wonder why had however many, 16 or 18 reviewed today and, we've got another batch coming in six months. Stop it. Like let's look at people as whole people.

Now that being said, I appreciate the comments you started with that said, the idea for this, you know, specifically looking at neck pain as opposed to you know, more (unintelligible) came from patients. I mean I'm interested to hear a little bit more about that.

However, I would say that we as a Committee, get to look at some of the big picture when it comes to things like functioning. And I have a question for NQF. Like what is our rule to look across the portfolio and think about how all the pieces come together?

And thinking about the burden of like, we keep carving people out. And then we wonder why they don't want to take the time to respond to any more theories, because they're getting 15.

Man: Yes, so actually that might be a good segue into a discussion of evidence. You know the questions for us is, is there at least one thing a provider can do to achieve meeting a change in the measure results? And given the evidence, do you (unintelligible) value the outcome of change?

And there was a lot of evidence presented but, kind of what do we think about that evidence. And then we can vote after we have a little bit of discussion on that.

Just as a point also, we don't have as much time to discuss this measure as the previous one. Technically we're supposed to be done around 10:30 this morning.

Man: So just to level set here, I think if we're using the term evidence, we should be talking about the quality and - the quantity and the quality. And I think those two are by no means related to each other. There's some association that quality does not equal quantity.

And I think if we're trying to evidence, we should ask about the quality of evidence. My evidence is an (unintelligible) of one, me. I've had a number of orthopedic and spine problems. And we have - and then have had, I don't know, probably 100 physical therapy solutions in the past five years, maybe more.

And it's helped but, you know to be fair, the group that I go to uses Photo. And so the nice thing is, it's really easy to collect. I mean it's just on an iPad and you do it.

But I have two questions for you. The first is, one thing that annoyed me was, as I'm looking through your question, there was no chance for me in the response, to say, it's not relevant to my life. For example, do you bend over to clean a bathtub? Well, I don't use the bathtub. And you know, what about soup on a shelf? (Unintelligible) I put it down here.

So some of the questions are challenging. But there was never a chance to say, not relevant to me. You have to choose all the way from one being really hard to, it's easy.

And how do you reconcile that relative to the score? Because I think it's introducing some bias to the - you know what I'm talking about.

(Deana Hays): I do. Thank you for the question. It's very relevant and it's an example of many issues that we grapple with in the development process. So what we're really interested in is the patient's perception of how they think it might be if they tried to do it. That's important.

Having said that, the beauty of Item Response Theory is that it will pick up if the patient is answering inconsistently. So if things don't match up, it doesn't negatively impact the score.

So the measurement - the measurement science is still good. However, that's a great example of lots of different - you called it an annoyance. There are a lot of different annoyances that we have to balance with the science.

Man: Yes, and that relates to my second question which is, as many times as I've filled that out, I don't think I ever had one conversation with a DPT (unintelligible) top, about how the score was being used and you know, how it related to my care. And that's pretty consistent.

So you know in one regard, are these data required by payers so that you can assess whether changed scores that aren't occurring would be subject to altering the prescribed benefit? You know, they usually give you a block of therapy, like 12 sessions or not. Is it at that level yet?

(Deana Hays): Yes, it is. And I'm sorry to hear that your particular experience was that (unintelligible). About your PROPM results.

The part of what the Photo system does is to train, to teach clinicians how to use outcomes in clinical care. And it's not something that we learn in our academic preparation. Usually, I think they're trying to do it in today's contemporary academic settings. But it's still not a skill that we easily pick up in our training.

So Photo does a lot to educate and teach clinicians just how to use the data to make sure that my perception is matching up with yours.

I could give you all kinds of examples. I use Photo as - I saw patients for 23 years, 15 of which I used Photo measures. And I could give you all kinds of examples as to how I thought I knew what was going on with that patient, until I saw their PRO results. And it caused me to communicate with my patient better.

Man: You must be a DPT. I guess my question is, are payers requiring the submission of Photo type responses to determine if deltas aren't occurring, show improvement. Patient reported outcomes aren't changing. Does that then translate into a change in the coverage decision of how many visits you might have in the future if it's not working for example? I'm just curious.

(Deana Hays): So far it has not been used punitively in that manner. The examples that we've provided in the submission were with respect to the MIPS Program and a couple of payer programs.

Man: Yes, I wouldn't use the word, punitive. I would use the word, economical or some mutual. Because we are trying to balance what we using as you know, physical therapy is under a bit of (unintelligible) for we use in some regards.

So I'm not trying to argue. I'm just trying to point that out. So the answer is, that's probably coming more.

(Deana Hays): You know we're encouraged, especially by the two payer programs that we featured in the submission are very encouraging because they show that when patient reported outcomes are asset focused, it helps shift out focus as providers, on communication with our patients, and quality of care, rather than how many widgets can I build.

And results so far suggest that when providers are called to attend to those results, they actually get patients that are (unintelligible).

Man: So yes, so that's some of the evidence. (Linda), did you have a question about evidence?

(Linda): I wanted to add that there's also accreditation consideration. And I know that Photo is used for of course requirements. So that it's not just payments that I think would affect its usability. I know that's not where we are yet. But also the accreditation requirements.

Man: (Linda) when you say, it's used, do you mean you check the box, we use Photo? Or is there some - how does…

(Linda): No, it's used for quality and performance improvement purposes.

Man: Okay.

(Linda): So lots of different body parts. But that it is how you - how certain therapists would identify areas of weakness within a system. And maybe look into specialty education for clinicians. So that's just how I've seen it used.

Man: Great. Thanks for that clarifying point. Any other questions about the evidence presented? (Ellen)?

(Ellen Sholts): So, one other thing that struck me, leading to the evidence section specifically is, you provided data showing that administering interim functional status assessments (unintelligible) care is associated with improved functional status.

Forgive me, but this feels circular to me. So if I've interpreting this right, like one of the processes of care, and specific one called out in the evidence base here, that providers can do like to improve their score on this measure, is to use the measure more often. Is that what you're saying?

(Deana Hays): Yes. And (Daniel), feel free to jump in. But what we think is happening there is that if a provider gets the PROM results sooner than later, as a physical therapist I'm more likely to tune into that.

If I wait until it's my patient's last visit - and oh by the way, do your Photo on your way out the door, and then I see what the patient reported, that I don't have any chance to follow up and say oh, I see you're still having trouble with such-and-such. Can we talk about that?

(Ellen Sholts): But so, what does that mean in terms of measured score here. Like how - so measuring sooner versus later. But are you using those measurements in the measured score itself to check (unintelligible) score?

(Daniel Dwisure): I'd like to address that (Deana) actually, if I may. Can you guys hear me?

Man: Yes, go ahead.

(Daniel Dwisure): Hello? Okay, thanks. So the challenge we have is to actually provide evidence that clinicians can do something in order to improve their scores. And the - I guess the obvious thing to think about is the treatment process.

Now Photo does not collect treatment processes as of now. Although Photo will start getting more treatment data as they integrate more with EMRs. But we have done another study during the last year, looking at the use of interim PROMs during the episode of care.

And obviously the clinical thought is that a clinician, just administrating the servers, would not be - would not always be beneficial for the clinical reasoning process done by the clinician and the patient.

But if they use the data more, that's why we want them to use PROMs. We want them to use the data, look at how the patient is doing, reconsider their treatments.

So the clinical thought is that we should be able to see a relationship between, use that as - at the discretion of therapists of the PROMs and the outcomes. And we wanted to look not only at frequency; how many PROMs they looked at during the episode of care, but also the timing.

Thinking that early timing will give more chance for modifications in the treatment plan during the episode of care.

If you looked at the methods we used, obviously there's some bias that could be involved in that related to the duration of episodes. And we use PSM propensity score matching, in order to match patients that did get interim - early interim surveys versus those that did not. But had the same chance of

getting an interim survey, depending on their characteristics as determined by the PSM.

So that - I think that evidence that supports the fact that there is something they can do. They can use the data and the patient reported outcomes in order to reconsider their treatments.

Man: Great. Thanks. All right, in the interest of time we need to start to move to voting. Let's vote on evidence.

Man: Evidence for 3461 is now open. Your choices are A, pass; B, no pass.

Man: Okay, we have 20. No, we have 21, yes.

Man: Voting is now closed and the Committee has chosen to pass evidence for 3461 with 17 for pass and four for, no pass.

Man: Great. Okay. So the next category is, Performance Gap. And I actually had a question about the Performance Gap numbers that were provided and what they mean.

In the clinician - well, individual clinicians as well as groups, the range and change was about the same. It was negative 0.5 with a big range of negative 14 to 22. What does that mean?

(Deana Hays): (Daniel), would you like to take that?

(Daniel Dwisure): So there were - yes. So those were the average or residuals by provider. I think the actual range was larger than the one you stated. We provided

additional information on the gaps at the two provider levels we tested - clinics and clinicians. And we did that by deciles of average residual.

So the residual would be the difference for a specific patient between the actual change score and the predicted change scores. So when you average these residuals and categorize providers by deciles of those residuals, we got a gap between the first and the tenth decile.

I think it was around minus seven to plus seven when we did that, using this method. And that was provided as an additional, or I think it was in the importance part of the submissions, that we submitted that.

Man: So this is residuals. For some reason I was like oh, so the average patient didn't get better? But that's not what you're saying. So okay, thanks.

(Daniel Dwisure): Right, right. Yes.

Man: Other questions about Gap?

Woman: So just to further clarify, like is a negative score better or a positive score? Like I think I would have to draw a diagram to figure that out so, why don't you make it easy for us.

(Daniel Dwisure): Yes, so the residual of zero for a specific patient would mean that their change score was the score predicted by the risk adjusted model. So if they receive the positive residuals, they would exceed their prediction. So that would be a good thing. And a negative would be a change score that's below the predicted change score.

Man: Okay, any other questions about Gap? Okay, let's vote on Gap.

Man: Performance Gap for 3461 voting is now open. A, high; B, moderate; C, low; D, insufficient.

Man: Looking for two more.

Man: Voting is now closed for Performance Gap for 3461. The Committee has voted to pass on Performance Gap with four for high; 13 for moderate; four for low; zero for insufficient.

Man: Okay. Now we move to Reliability and Validity of the Scientific Methods Panel. Passed it on reliability consensus was not reached for Validity, with one low and one insufficient. But all votes for Reliability were moderate.

(Sam Stopey): So just a point of clarification related to the Validity voting. So the Scientific Methods Panel had a kind of specific concern related to this measure. And they were looking for a specific test.

Now the developers subsequently provided that test. But because it wasn't present during the deliberations by the Scientific Methods Panel, they noted that they would have passed the measure, but necessarily achieved a consensus not reached, because they didn't have that specific test.

Subsequently the developed provided it. The staff evaluated the measure as high, for validity. But because it was CNR with the Methods Panel, you do not have the option to accept the Methods Panel evaluation.

Man: Okay. So comments from the discussants on Reliability and Validity and then - no, I didn't see it. There it is. Okay. Okay so no input from the discussants on Reliability and Validity? Okay.

Right. Again, yes, additional data were provided. The Scientific Methods Panel I guess, felt better about that. (Sherry), anything jump to mind from your standpoint?

(Sherry): Yes, well I missed that scientific - I wondered why these weren't sounding familiar.

Man: You seemed quiet, right.

(Sherry): I was not at that meeting. There was a death in the family. So, but I do have a question because - a couple of quick questions for the developer about Reliability. So is it Reliability we're still on or, is it both?

Man: Well, we can kind of discuss both. We're still on Reliability technically, because we haven't voted on it yet so, yes.

(Sherry): Okay. So the Reliability signal to noise formula you gave, didn't include in the denominator the variation - at least I didn't find it, the variation within patient across items in your measure.

So, you know, how do you think that would influence the data you presented? So that's question number one.

(Daniel Dwisure): So when we - yes, so when we presented Reliability we did - so I'd have to get a clarification from on what you mean. Because we did look at the ratio between the different levels of variance within and between providers.

And you're asking about the reliability at the provider level - at the score level, right? Am I getting that right?

(Sherry): Yes. What you usually do when you have multi-item measures for a patient, and the error term includes patient variation across items within patients, and plus the variation across patients within provider. And then plus the variation between providers.

So the denominator needs to include two error terms. And I didn't see it in the formula that you provided. And I just assumed that…

((Crosstalk))

(Daniel Dwisure): Yes, so the denominator included the variance of providers to providers and the specific provider error within provider.

(Sherry): No, no it's not the provider. It's within patient across items that I was looking for. Because you have to include that when you're thinking about the error - the standard error of measurement that there's two error terms at the provider level.

There is patients reporting a variation in patient level data across the items for each patient. And then there's variation within provider across their patients for the score. So the…

((Crosstalk))

(Daniel Dwisure): So we followed the…

(Sherry): I know. See that's…

((Crosstalk))

(Sherry): Yes, sorry. I'm interrupting you because that's my fault. Because that's one of the Scientific Methods Panel's considerations now. And it probably wasn't the guidance. (Lisa), is that right? It wasn't the guidance the measurers developers were being given, right. So (Sam), you want to…

(Sam Stopey): Sure. So (Sherry) you were specifically excluded from this particular group, not for any sort of spite, but because we knew you'd be on this one so, we can't double-dip.

But this was carefully evaluated by (Z.Q.) and the process, and he gave direct feedback to the developer on what sort of calculations they should perform.

(Sherry): Yes.

(Sam Stopey): Specific to the signal to the noise analysis. So they followed instructions according to (Z.Q.)'s.

(Sherry): Okay. So that - then we're looking at that now and that may change in the future. So that's one of the things.

But the other - one other question I have or two other questions is that in hierarchical or nested designs, and you've got patients within clinics. Clinicians within clinics, the variance inflation factor has you thinking about the number of clinicians you need per clinic to get a stable estimate of the clinic's performance. And then you've got some confounding in the small clinics, small groups. So the clinician is the clinic.

Have you thought about any about how that sampling process would go? Because if you've got low - you've got the confounding. That is a big problem. And I'm assuming that you've thought about that.

And then, how many clinicians within clinic you need to get a stable estimate of the clinic's performance. Is that how you're seeing this being used?

(Daniel Dwisure): Well, yes. So for the clinic level we just differentiated between large and small clinics. Because the reality in the field is that some clinics may have just two or three clinicians. And when that was the case we looked at it more as if it was on a clinician level.

So we took into consideration the thresholds that we used for the Reliability testing. We took the number of clinicians within clinic as a consideration for those thresholds.

So I'm not sure it answers exactly what you're asking but, it may. Let me know.

(Sherry): Yes, that might be one of those things going forward. But my final question is, the root mean square of approximation. And you got some - and the way I read it, at .16, which isn't great, could you clarify how you handle that?

(Daniel Dwisure): Right. So during the development process, and there was a question about the - about that paper (unintelligible) in the introduction comments, so that was published in a Peer Review Journal, in the Journal of Orthopedic - JOSP journal of Orthopedic Physical Therapy, which is a very high level Peer Review Journal within the Orthopedic Physical Therapy literature.

So the process of developing the final item banded from the item pool, the process is a multifaceted process, as obviously you know.

And one of those steps goes through a factor analysis, obviously. And during that factor analysis the results were those that you stated, with (unintelligible) being about 1.6, after which a few other items were deleted.

And what we found out in that paper as the three factor solution would improve the statistics. But then there were also clinical considerations about the impact of having three factors within this measure.

The next process after looking at those FIT statistics, went through the IRT development and the IRT FIT statistics. So a few of the items were deleted after we saw those results in the previous process, which was the exploratory and the factor analysis.

So the final FIT statistics are probably slightly better than what you saw at that stage. Because it's just one stage of the - one part of the development process.

But again, because one of the factors that seemed to stick out as a separate factor is related to sleeping items. And some of those items were later deleted. And that's explained in the manuscript, because of low FIT statistics with the rating scale IRG model. So the final FIT statistics are probably better than what we reported there, because they were not final.

Man: Any other Reliability concerns? Let's vote on Reliability. The question is, do you accept the Scientific Method's Panel rating for Reliability, which was, moderate? We need one more. We need no more. Okay.

Man: With voting closed the Committee has decided to uphold the Scientific Methods Panel rating for Reliability of moderate with, 21 for yes; zero for no.

Man: Okay. And then for Validity we have to vote on the individual elements since there wasn't a consensus. So any further concerns about validity? (Don)?

(Don Casey): Just briefly, you know, having been a guy who had more than one thing going on. And taking Photo specific questionnaires relative to the one thing when it's related to the other thing, without getting into detail, how does that - I'm just not clear about how you take into account how these, for example, answering some of these functional impairment questions might relate to the fact, it actually could be due to the other thing. So I guess.

Man: So I guess (unintelligible) I would wonder what proportion of the patients in the test sample had multiple conditions that might of…

(Don Casey): Well, let's say I was in an auto accident and I fractured my leg and hurt and twisted my neck. So I'm just…

Man: Or hurt your back or you know, the other thing that you said.

((Crosstalk))

(Deana Hays): So…

(Daniel Dwisure): Go ahead (Deana).

(Deana Hays): Go ahead (Daniel).

(Daniel Dwisure): So the great majority of patients that are asked, when they come up to the kiosk or use their iPads to - the iPads to respond to the survey, so one of the first screens, while the set up - the staff sets up the surveys asking the patient to decide on their main body area that they're seeking treatment for.

The great majority - almost all patients have no problem selecting a body part. Because they usually do have one main area that bothers them.

Now your point is well taken because that's not always the case. And not only that other body parts affected might impact their scores, other - comorbidities in fact, could do that as well.

So if a patient comes in and they have a knee injury and were asking about physical functions they, I don't know, they have (unintelligible) or other comorbidities, it would be difficult for them to differentiate what part of their functional challenges or their perceived functional challenges are related only to the knee or to their (unintelligible).

So it's a similar type of question. And that's a very common question that clinicians need to address. But that's - as we see it, it's not really the role of measure to differentiate that. It's the clinician that needs to take the perception of the patient or their functional ability and cross-check that with the overall condition. Whether it's other body parts or much more often, other comorbidities. We don't…

((Crosstalk))

(Don Casey): I'm sorry, I don't mean to interrupt. But in the interest of time I'll just say, I wasn't given a choice. I was told, since I was here, physical therapy approved

for X, and I had Y, too, I was handed the tablet and it was, fill it out for X, I'm just saying.

((Crosstalk))

(Don Casey): So all I'm saying is there's some - we don't need to argue it. I just want to point out that there's still some concerns about (unintelligible).

((Crosstalk))

(Daniel Dwisure): So that's a - yes, that's a - well (Deana) might add something but, I think that's a bad example of how a PIRM needs to be set up for a patient.

The patient is the one that needs to decide well yes, my main issue or this or that. And not set up because of something written in a diagnosis on a referral.

So again, I'm sorry about the experience you had. But that's not the way it's being taught.

(Don Casey): It's not intended but it is common, all I'm saying.

Man: Well, and I think it bears note for use and usability, especially the next time we look at this. Okay, anything else on Validity? Otherwise, let's vote. Okay, let's vote.

Man: Voting on Validity for 3461 is now open. Your options are A, high; B, moderate; C, low; or D, insufficient.

Man: Now we're 21. Good.

Man: Voting for Validity for 3461 is closed. Validity passes with one vote for high, 15 votes for moderate, five votes for low, and zero for insufficient.

Man: The other Validity thing was right? Oh, Feasibility is next? Okay, good. So, there's no more subgroups, okay. Very good.

Okay, Feasibility, any of the discussants. I think it was mostly usability and use people had concerns about. Any Feasibility concerns. It seems pretty straightforward. Oh, (Ellen).

(Ellen Sholts): So, I would just reiterate my earlier comment that, you know, when you start having these very narrow measures and then you put that together, it can become a burden on the patient. So even if this one item isn't burdensome.

Man: Additive feasibility?

(Ellen Sholts): Yes, additive feasibility definitely could be a concern.

Man: One - there was one of the pre-evaluation comments that asked for more clarity on the time and cost clinicians may encounter. Does that sound familiar to anyone who'd like to talk more about it? (Sherry)?

(Sherry): Sure. I mean it was just as a matter of how it's administered. So through the - within the practice and what that means for physicians that are in the practice.

(Ellen Sholts): Is this about cost?

(Deana Hays): Both cost and - so timewise it's very efficient, five minutes or less. If I could circle back. For the physician - oh, no time for the physician. Depending on -

well, it depends on how you administer. So I want to be clear that the tools are free. They're available, free for use.

The patient reported outcome measure and the risk adjustment model are free for use. There is additionally, a free Web site where the provider can sit the patient down so that the Web site administers the CAP version of the PRO and the risk adjustment questions for the patient and generates the scores for the provider. So if they use that for free, the administration and the scoring is done for the provider.

There are additional services that can be provided by Photo at very low cost to generate reporting to the clinician within, you know, in real-time so they can use it at the bedside with the patient. A number of other services to help the data really come alive and thrive in your practice so you can improve your quality.

There's the low level service if you have, say a practice with three therapists, the first year it's $1000. The second year it goes down to $850. So it ends up being a very economically feasible addition. For what it's worth, almost everyone that subscribed to the Photo services chooses the higher level because they want everything. And it's still very economically feasible.

(Sherry): And can I ask related to that, in the comment that came from the public and how I think it's not perceived as being interoperability, due to the lack of I guess, a LOINC code, to have it be you know, layered into the patient's record. Is that accurate, based on the public comment?

(Deana Hays): So the interoperability is something that we're very concerned about. A lot has gone into integrations with multiple electronic health records, again to

reduce patient and provider burden and interoperability, standardization of data elements such as the LOINC.

LOINC in particular we have not looked at. That's the first time I've heard anyone mention the LOINC codes. But interoperability and common data elements and such, is of great importance to us.

Man: Okay, anything else on…

(Sherry): The computer adapted testing means that you're reducing the patient burden by adapting it on different basis of difficulty, right? So that you would not necessarily ask all patients all questions. So that would add to the sort of feasibility of your measure, right?

(Deana Hays): We believe it does. And that's why we've put so many resources of just providing the CAP version. Because through the CAD and the IRT, you're able to harness the power of a large item bank without the patient having to answer all the questions.

(Sherry): Right. And so - and just one thing about the difficulty, you know, in small little body parts versus the whole (unintelligible). The problem is that then when you go to attribution, the clinicians say, wait a minute. Don't hold me accountable for that because that's somebody else's job. That's a specialist's job. I'm doing whatever.

For primary care, okay. But for some of these very discreet problems you get the issue of attribution being problematic. And I just think that that's, you know, one of the things that you're struggling with. We haven't talked about attribution which is kind of a rabbit hole. But, that is an issue.

(Deana Hays): So, (unintelligible) wants to jump in on this too. But I will say like if many, many more of our measures were holistic and not necessarily setting a provider specific, we would have a more holistic healthcare system that probably would have better care coordination and better connections, and few more silos. That's my view.

So measures could drive that change. And that's very much what we hear from patients and families that they want is like, I want one system, not 15 systems that I have to navigate.

(Sam Stopey): I think the attribution target though is the therapy centers and the therapists. Not the doctors and the prescribers. I mean it could be related but, I think it's really targeted at the therapy providers which as you know, is problematic from the standpoint of concern about overuse.

Man: Okay, great. Those are important feasibility questions. Anything else related to feasibility, we should vote on that? Okay.

(Daniel Dwisure): Could I maybe just add one short comment about the use of different body parts - measures for different body parts versus more normal measures? So just a quick note as a follow-up to what (Deana) said.

So specifically for the neck, we came from a more global measure with having a lot of complaints from patients and their clinicians about difficulties relating to specific items. Most of those complaints kind of went away when the items were worded a little bit more specific to their own problem.

However, the function that is usually asked about, with the perception of the difficulty level of a specific function is, many times global. However, we're

aware of the concern you're raising, which I think is very accurate to raise as a concern.

And just as an example, Photo has three separate measures for lower extremity, which we are now looking into merging into just one - a few measures for foot and ankle. And for the sake the advantages of more global measures, we're looking into merging those.

So we're aware of the pros and cons of those two issues. And I think it's a great discussion point. Thank you for bringing it up.

Man: Okay, let's vote on Feasibility and then we'll have a very brief discussion on anything else with Usability and Use.

((Crosstalk))

Man: It will be there. It's there now.

((Crosstalk))

Man: Voting for Feasibility is now open for 3461. You can vote for A, high; B, moderate; C, low; or D, insufficient.

Man: One more. No more; good.

Man: Voting for Feasibility for 3461 is closed. Feasibility passes with 15 votes for moderate and six votes for low.

Man: Okay.

(Sam Stopey): But a point of clarification (Chris). In this case the measure developer is the owner. And obviously he's going to benefit from NQF endorsement. I don't know how we process that. I mean we have other people coming in who are measure developers as an independent entity.

But I just want to make that point here because I think it is an important issue to consider.

Man: Yes. And, anything from more senior NQF staff about that? Just, do we need to note it?

(Don Casey): May I ask, that is a - we see that in other measures. Is that a relevant factor or not? That's a real question.

Man: It happens.

Man: I mean should we consider that in our deliberations or not?

(Ellen Sholts): And I would like clarification if it's in the public domain.

(Deana Hays): Yes, so it is free for use by the public. So, and that's detailed in the submission. The components are available free for use.

I'd also like to mention you know, regardless of who develops the measure, there has to be funding. Whether it's grant funded or what it is, you want measures that are owned and stewarded.

Measure - you know, grant funding runs out, too. Our friends at Northwestern University are facing very dire circumstances because the NIH funding has

run out. And we very much want them to succeed. They're having to go to a business model.

So you want something that's owned and stewarded and taken care of, not just put out there. But that's why everyone has to publish and in peer review journals, and go before a NQF rigorous panel, so that any risk of bias or other concerns are adequately vetted.

(Don Casey): But with all due respect, it's free. But you do license it to large organizations like Athletico. And you do make profit on the technology behind it - because it does take technology to do.

So let's not argue the point. I'm just raising this as a question for NQF, and how we go about doing this in the future, I think in my opinion, is an important issue.

Man: I think it's important to just put it out there, right. I mean we have had other measures that were almost ready for prime time but then, couldn't make it across the finish line because funding was lost. And so yes, it's a tough thing.

Okay, Use and Usability. Use is first, right? Okay there were one or two concerns in each category and (Sherry), you had voiced a couple at the beginning.

(Sherry): Looking exactly at what I said. But I'm relooking at, under the accountability component of it, you know I know it's in a QCDR which is one thing because those can be used for quality improvement purposes within a practice.

You know, but it mentions that CMS does plan to make all measures under MIPS available for public reporting. I guess if this is a measure that's

intended to be at the center or group level or practice level, MIPS is a physician specific program. The measures are tied to an individual clinician. So how would that be reconciled for the purposes of this measure?

There are group reporting that you can do but, the measures themselves also are specifically tied to ten MPI.

((Crosstalk))

(Sherry): Unless it stays only within the QCDR, I supposed. Is it looking - are you looking to keep it only within the QCDR environment, or have it be an independent measure which would be a ten MPI?

(Deana Hays): Thank you. I understand. We - the QCDR measure is a faster path. So we first submitted it as a QCDR measure so that it could be available to lots of providers who really wanted it and needed it as quickly as possible.

But at the same time, it went through the measures under consideration process to become a MIPS clinical quality measure.

And it was passed by the MAP Committee in December, pending NQF endorsement.

Man: Okay. Any other use questions? Let's vote on use.

Man: Use for 3461 is now open. You can vote for A, pass; B, no pass.

Man: One more. One more vote. Everybody voted? Do we just go with 20? Yes, let's go with 20. No, we got one. Okay.

Man: Voting for use for 3461 is now closed. Use passes with 18 for pass and three for no pass.

Man: Okay, great. And then (Sherry) you had a question or two. There were a couple of questions about usability. The pre-evaluation comments were all positive but, just a couple of things, if anything hasn't already been discussed enough. Otherwise we can just say it's been discussed. (Terry).

(Terry): Just a quick general comment about usability and the need for linking data elements to standardized vocabulary, the LOINC reference that (Sherry) made.

This is becoming increasingly more important, particularly as EHRs for the repository or more and more information, they can't talk to each other without a shared vocabulary.

So in future considerations of measures that are electronically measured or any of them, they should all really be linked to something like LOINC.

Man: That's good. From a NQF standpoint, what's going on in terms of trying to get consistency? Anything yet? Consistency of language and terminology and so that the measures are consistent, what's going on. You know, whether it's EHRs, anything yet? Or is that a new frontier we need to talk about?

(Lisa): It is part of our discussion. So we discuss a lot of things here, including trying to make sure that we're providing consistent guidance to the field.

We don't have anything yet but, it is part of the work that we're doing on ECQMs that's being led by our lead, (Chris Mollette) and (Katie Streeter). Where this is an issue that we're bringing into the CCSEC again, in October.

Man: Good. Thanks (Terry) for raining that. (Sherry), one more point.

(Sherry): Yes, I mean I think my comment on this is similar - oh, I'm sorry (Felicia), you had your card up. Sorry about that. Was similar to the previous one in terms of - and actually what (Don) brought up related to what the physicians or clinicians or PTs are able to do with the information that's received. And whether it is - and I know you mentioned providing education.

I don't know if that's provided at cost as well. But in order for the clinicians to be able to act on the data received in an appropriate and effective way.

Man: Yes, that's a good point. We were discussing the same thing with the last measure at the break. (Felicia), did you have a point?

(Felicia): I was just noting that - (Lisa) addressed it. That this isn't submitted as an EQCM. And so it doesn't - there's no requirement to make it more incompatible or anything like that.

Man: Great. Thanks. Okay, let's vote on usability.

Man: Usability for 3461 is now open. You can vote for A, high; B, moderate; C low; or D, insufficient. Voting is closed. Usability passes with 13 votes for moderate, seven votes for low, and one vote for insufficient.

Man: Yes, we need the percentages for that. Sixty-two, okay, okay.

Woman: So NQF 3461 does pass with 62%?

Man: All right, overall suitability.

Man: Overall suitability for 3461 is now open. You can vote for A, yes or B, no.

Man: Two more? We're good.

Man: Voting for overall suitability for 3461 is now closed. The Committee recommends NQF 3461 for endorsement with 14 votes for yes and seven votes for no.

Man: Great. Thanks everybody. We've gone pretty over but, I think it was a good discussion. Any recommendations for continuing or are we going to keep plowing right along? Okay. Three second (unintelligible).

Woman: We're going to move on to two measures that a number of us on this committee know well, from the first time they came around. And we're going to depart a little bit from the usual format in that the developer is going to address the next two, together. And here they are. Now would you like to introduce yourselves? And just for the record we're addressing 2286 and 2321.

Kathy Dann: Hello. I'm Executive Director with UDSMR. On behalf of UDS and the subscribers, we're pleased to be here today representing NQF endorsed functional quality measures for inpatient rehab facility. They're the same ones that have been endorsed for the skilled nursing facilities and the long-term acute care hospitals. Thereby able to facility impact that. Excuse me.

These measures were developed out of extensive research and more than 30 years' experience with functional measure - with the functional measure known as the FIM instrument.

There are more than 3500 articles related to the FIM in just the past ten years. It's being used in 24 countries for quality purposes. And it's an additional 31 more countries for quality improvement evaluation and research.

To clear the air on the proprietariness of the measures, CMS already has a royalty-free license in perpetuity, to utilize the FIM instrument which would include these measures in any setting.

We have offered a number of (unintelligible) in the public domain, if that's what's needed for the greater good. And I want to clarify it. We don't get paid for usage of the instrument, only for our services in conjunction with it.

CMS is just one user of the instrument. Multiple countries, venues, and others still have a need for its usability after 10-1-19.

A recent survey of a subset of our subscribers and our subscriber's represent 83% of all inpatient rehab facilities, indicate that 50% of them are likely to continue to utilize the FIM or portions of it in their programs for quality purposes for the following reasons.

One, the desire to monitor cognitive elements. Two, it's required by some non-Medicare payers. Three, it's used for carve and joint commission accreditation. Four, a lack of trust with new unproven CMS mandated measures. And five, a case management in general. And these are directly out of the survey that we did with our customers.

I mention this a little bit as an indicator that we do get feedback. And 30% of our customers have been with us 20 years or more. So these are long-term inpatient rehab facilities.

We thank the Quality Forum - the National Quality Forum and the Patient Experience and Function Committee for their time and consideration of our measures. And we welcome the opportunity to address any outstanding questions related to our university affiliation, non-for-profit organization. The measures (unintelligible).

I have with me Dr. (DexAnne Aquila), a Physiatrist with particular chief and functional assessment. And our Director of Research, Dr. Paulette Niewczyk, can't be here today because their flight was cancelled. She was supposed to be here. But she's on the line and stands ready to talk about the characteristics pertinent to the committee's interest, including the measured development process and the measure testing results that we have provided. I'm going to turn this over to Paulette to take this forward.

Paulette Niewczyk: Hello?

Kathy Dann: We can hear you.

Paulette Niewczyk: Okay, thank you. We have two measures, and I'm speaking generally about both of those together so it's a little different than what we've been doing thus far. That includes the self-care measure and the mobility measure.

Both of these measures have demonstrated reliability and validity. They have a high, overall consistency. The ability to capture significant functional gains, high discriminative capabilities, and are predictive of change in function, as well as, likelihood of patient discharge from rehab to the community setting, in addition to length of stay.

The self-care measure is eight items. There are six physical and two cognitive. And the two cognitive items, you know, going off of the last conversation, are really meant to make the self-care measure more dissolved.

Where completely independent self-care requires both physical and cognitive functioning, unlike mobility which is all physical. The mobility measure includes four items.

Both measures are intended for use on all persons aged 18 and older that are treated in an inpatient post-acute care setting. They're for all impairment groups and all payer sources.

They're both rated on a one through seven level scale, where higher is better or seven would equal complete independence. One would equal complete dependence or helper assistance needed.

I understand there were some concerns related to usability. In particular, plans after October of 2019. I hope that Kathy Dann had address that. CMS of course has made a decision not to include the self-care mobility items or the FIM instrument moving forward and, into the IRV PI. But the IRV PI tool is meant for payment purposes.

Our measures are meant for quality and patient outcome purposes. So really that holds no bearing. It doesn't have any effect. We still have a need to measure patient outcomes. And we will still have measures available to fill that need.

I also had sent, and I trust that the Committee has received some additional data, on June 6, related to changes in facility performance over time. Both the

self-care and mobility measures demonstrated improvements over time from the years 2015 to 2018.

And related to evidence, I don't think it was very clear in our measure submissions, but post-acute care and medical rehabilitation is the service that is anticipated to improve patient function. The measures should be capturing that change in function. But the service would really be clear delivered at an inpatient rehab facility or a skilled nursing facility or, a long-term acute facility.

And both the self-care and mobility measures were able to capture patient change as the admission items were significantly predictive of patient discharge to community, which is a goal of inpatient medical rehab. They were predictive of inpatient length of stay, as well as, patient change in function.

So at this point I'd be happy to answer any questions related to the measures. Thank you.

Woman: Thank you. Before we start addressing our questions to our developers, I do want to make the point that as we all know from looking at the agenda, we are going to come back and discuss these measures again this afternoon, in the context of competing with two other measures.

We are not having that discussion now. What we are doing now is deciding whether or not each of these measures merits continued recommendation from this committee for endorsement. So with that (Ellen), you're our lead discussant on self-care.

(Ellen Sholts): Okay, and just to be clear, we're going to focus on self-care first before we go to the second measure.

Woman: To mobility.

(Ellen Sholts): Okay. So just looking at the evidence as a brief recap, so the developers provide a logic model reflecting you know, what they just described in terms of having access to post-acute care relating to improved function.

They did not provide evidence around a specific structure process or intervention to be connected to the measure. And so that was something that I had a question about. I'd like to hear a little bit more from the developers about that.

They do provide evidence that the self-care measure correlates to positive outcomes. The outcomes specifically that they list is that, if I understand this correctly, that the score on this instrument correlates to the score on the larger FIM instrument.

So one question I had is, like I mean again, isn't that sort of circular? I mean I would expect a strong correlation. You're taking a subset of your instrument and correlating it to the whole instrument.

To me that doesn't answer the broader question of, you know, what is the broader outcome or, you know, what other source of outcomes is this correlating with?

They also provide some information about how the outcome correlates with patient discharge to the community. So that was highly, specifically significant change in function and self-care items. Also, for items retained in

the model. And they also show that there was a significant correlation with patient length of stay.

However, it's not specified which direction that correlation was in so, this s another question I have for the developers. Did you see, you know, higher functioning…

((Crosstalk))

Paulette Niewczyk: Yes, so greater patient change was associated with the increased likelihood of discharge to community for instance. In terms of length stay, we actually see you know, the inverse.

So, you know, if it's a larger change then we typically will see the length of stay will tend to be also larger, before more treatment was warranted. And then for -- I have three of them -- and then of course for the extent, so it would be the admission self-care score was also predictive of how much change you could expect.

(Ellen Sholts): Can you say a little bit more about that? So for example, if their admission self-care score was higher, so they came in with higher functioning, then what is the effect that you're seeing?

Paulette Niewczyk: So basically what we're seeing is, we have a narrow window, right. So most patients that are treated in an inpatient rehab facility have the ability to withstand three hours of therapy every day or most days of the week.

So those that would be of the lowest level of function often are not going to an inpatient rehab facility. Now I'm speaking in generalizations. There's always

exceptions to that rule. But you need to be able to withstand three hours of intensive, multidisciplinary therapy.

So they tend to you know, cluster in a certain range in terms of where their function is at admission.

What we find is those that have, you know, slightly higher level of functioning. So not higher whereby they're coming in at all sevens on each of those items within the self-care, but certainly they're not coming in at all one's for instance.

But those that tend to be more able to withstand and tolerate three hours a day, tend to elicit the largest extent of change when it comes to discharge scores.

Woman: Could I make a sort of a clinical observation on that from a physician's point of view. If you think about inpatient care and in this measure that's what we're talking about, there's not a specific length of stay that's good or bad.

What you're trying to do is to get a patient to a level of function where they no longer need inpatient care. And so, I mean that's kind of the goal. So the lower a person comes in, the more change needs to be achieved in order to get them to the point where they can do well outside of a new (unintelligible) setting.

So it's not surprising to say a person who comes in at a lower score is going to make more gains. Not really that they are a better or worse candidate for rehab. It's just a practical thing. You're trying to get them to the point where they can go to the next level of care. Not that rehabilitation ends at time of inpatient discharge. Just changes to different venues. I don't know if that helped at all.

(Ellen Sholts): I think we're mixing up a couple of different things here. So we're talking about evidence. So I'm looking for a structure or process of care that can improve this particular outcome. Improve the overall change in self-care function, right.

So one of the pieces of evidence that you presented is patient length of stay. My question is, is a longer length of stay associated with a better outcome. So that would be like a larger change in function.

Or is a shorter length of stay associated with a larger change? Because to me that makes a difference of whether that length of stay is an indicator of the level of functioning upon arrival. Or whether it's, you know, like having a longer period of time in which someone is accessing all of that therapy.

((Crosstalk))

Paulette Niewczyk: Yes, I think I understand where you're getting at. You're looking at something like a trajectory of recovery, right?

So length of stay is a little tricky because length of stay is often anticipated when a patient is admitted. So some of that is outside of the scope of what you know, a measure would be able to capture if there's particular payer and they're allocated a certain number of days in rehab.

That may be independent of their change or their functional improvement or, what their level of functioning is at admission. Does that make sense?

(Ellen Sholts): Yes. I just know that this is the evidence you presented.

Paulette Niewczyk: It is. It is, absolutely. But the goal of inpatient rehab is to get the patient - number one, to improve the patient's level of functioning. Number two is to get the patient back to a community based setting, or where they were prior to you know, what led them to needing this type of inpatient care.

So those values, I tend to attribute more to that process what you're talking about. And that process is the, you know, the healthcare service. It is that, you know, multidisciplinary rehabilitation. It is the post-acute inpatient care.

(Ellen Sholts): So then that sounds like this is really about access to care. But this measure is going to be used to compare different inpatient rehab facilities, right. So what is it that one rehab facility can do different from another, that is going to lead to improvement on this outcome? That's what we need to see for evidence.

Paulette Niewczyk: So that's actually what I did provide with the data that I submitted for 6-6, which is that facility change over time.

(Ellen Sholts): Could I give an example of how it's used? So literally sitting in a patient conference that occurs once a week where the whole rehab team comes together, led by a physician with nurses, physical therapists, occupational therapists, (unintelligible); whoever needs to be in the room is there.

And they can look at these numbers and see what goal are we driving at? What SIM level do we need? What functional level do we need for this patient to be ready to be discharged?

And you can graph this out literally, these numbers on a chart, to see where a patient is making progress and, where they're not.

And back to your earlier point, it is perfectly common sense that the more gains your team needs to make your patient achieve, the longer it will probably take you in-house, to do that.

There's another metric that you can use. You can take what we call, length of stay efficiency, where you look at how much gain - functional gain is achieved, and how long did it take you to achieve it. Because you're also trying to figure out ways, can we do it more efficiently and not have to keep the patient in-house.

But literally then the team sits down and says, oh my, goodness. It looks like we're not as far on advancing the patient in self-care and toileting as we need for them to go home. Because a patient's ability to manage toileting self-care is one of the primary determinants of whether they're return into the community or whether they'll spend the rest of their days in an institutional setting.

So literally the team can sit down say, we're not making the progress we need to make for functional level for discharge. How do we shovel the coal on? Let's spend some more time with an occupational therapist. Let's get a piece of assistive equipment. Let's bring in somebody that's ore engaged with the patient and can get the patient more motivated to participate.

All of those are the processes, interventions and structures that then are driven by these measurements. So it helps you know what you need to do and where you need to focus.

Woman: Just to quickly interject, the data and additional information that Paulette referred to is in the Competing Measures memo that we sent out on June 6. So if you go to the Committee SharePoint home page, it's up at the top under,

Materials, Competing Measures Memo. And there's an appendix in there. We sent that out a couple of weeks ago.

Woman: Is it your view it should come up - it should be incorporated in the discussion here?

Woman: Yes.

Man: That's part of the evidence.

Woman: The additional information the Committee requested during the orientation call is in that document.

(Ellen Sholts): Okay. So I think many of us are going to look at that. I've used a lot of time so I want to turn it over to my co-discussants.

Woman: (Don)?

(Don Casey): Thank you. I actually spent a lot of time on this. There are just a couple of things here. Again, mystery evidence is here. So I was able to find a large number of studies, like 900, between the last reference that was used in the submission which is, I think, about 14 or 15, including some from 19, using that (unintelligible) strategy and FIM scoring inpatient rehab.

So just begs the question about how measure developers are not presenting evidence, in my opinion, in a way that helps us really make a better decision. So it's not a complaint as much as it is an observation that this is a big gap. So I want to put that into play. And maybe that helps the team here.

I will say, and again not getting into the conversation (Lee), that you don't want me in, I will still say that I called their results that they presented in the Competing Measure Memo, good. In the sense that the IRV with overall scores of greater than 75%, had actually risen from 28 to 39% over the most recent four year period.

And then the other thing is, and I think this was (Ellen), you were sort of getting at this a bit more. Healthcare disparities are not really called out in any elegant way. And if you think about social determinants of health, sociodemographic differences, access to inpatient rehab facilities. And that is not just in the city, that's in rural areas that are far away from home.

There isn't any mindset about how to figure out how that impacts the selection and measure of these patients. But logically, having done a lot of work on an inpatient rehab facility for geriatrics (unintelligible) during my clinical career, that's a big issue and it impacts a whole lot of things in terms of the outcomes.

And I just again, want to emphasize that I don't think social risk factor variables in the form that we filled out, is well specified. And then let me just see if I missed anything. Well lucky for you, my computer is frozen. Thanks.

(Ellen Sholts): I just have one quick question or thought. The evidence that you presented suggest that 37 facilities were in the bottom quartile. And that's in the performance score thing. And 22 were in the top quartile. That's 7% of your whole sample at the extreme. Everybody else is tightly knotted in the middle.

So what would have been helpful to use to understand is, what's the evidence that there is a lot of variation at the facility level in that inter-quartile range, number one. And secondly, it's hard to interpret that what a clinically meaningful difference is, without some evidence for that.

And that would have been helped by some discussion about you know, what is the variation that the facility - that between plus within facility variability. So did you do analyses of that type or no?

Paulette Niewczyk: Okay, so I - the table that's below the facility level table is actually showing what their facility mean change and their patient level scores are.

So here you're actually seeing, you know, those that were in the first quartile or the below 25%, what their mean change in self-care score would be. That's 3.75 for 2018 for instance. The (unintelligible) would be you know, eleven. The 75th would be almost 16. And then the above at 23.

So you see, there's a pretty good spread actually, in terms of where their scores would be centered around.

We're not surprised that many are falling in that 50th percentile. So they're giving, you know, good care. Maybe not excellent care. Reaching that 75th percentile is going to be much more of a challenge but, it's certainly attainable.

And I think it's a good thing that there aren't a whole lot of facilities that are in that, you know, first quartile. So in terms of where their ranking is, there is change that's happening over time. There are some facilities that are going from, you know, the lowest to the next year, bumping up into that 50th percentile.

But we're also seeing some you know, top performers are well. But for the most part it's going to be centered around the average of the 50th percentile. It's where the bulk is going to be.

(Ellen Sholts): I'll say from a clinical viewpoint and a practice standpoint, institutions who use these measures flight and claw for every percentile improvement they can achieve in comparison.

And then they become very I think, clinically, appropriately creative in trying to figure out, how do you do that? How do I get 1% better? Do I send my physical therapist to a special training class? Do I do this, that, or the other.

One of the things you can do, because you can go down to individual patient level and care giver level, you can look and see which particular caregiver, therapist, physician with a certain diagnosis of patient is achieving better functional gains. And ask that person to talk to their colleagues. What do you think you're doing that may be effective?

So these measurements, you know, you could put a gun to my head and I could talk a little bit of rash analysis if I had to. But that's not my pursuit. But what I can talk about is, do clinicians feel that these measures are sufficiently accurate in measuring a patient's functional performance, which translates to their ability to go home, the amount of care they're going to need from family, friends, paid health, all of those things.

Do we think it's efficiently accurate to face and drive clinical decision-making on it? The answer is, absolutely yes, with great degree of healthy competition. People are trying to get their patients better and to have it be demonstrated in a measurement with this level of authority, reliability, validity.

Paulette Niewczyk: I provided, you know, a very consolidated version here. But this entire thing could be replicated at the facility level. Whereby they see what their facility average change in self-care score was for 2015 and 2017.

(Ellen Sholts): Yes, yes.

Paulette Niewczyk: So even if they're at the 50th, they may be making some substantial changes where they're going from an average of 11 to 12 and now 13. And they're inching their way up to that next quartile.

(Ellen Sholts): See, that's what we don't have. And at the patient level it's fine. But if it's to be used at the facility level, what we don't have is a signal to noise analysis. Because the top 3%, the odds are, you've got ceiling effects and there's going to be a lot of signal to noise problems.

And the bottom three - the bottom 4%, you're going to get regression of mean over time. So you know we don't - I don't see us having quite enough data to answer that question at the facility level.

So I'm listening to not, you've done it but we don't have it. Is that what I'm understanding?

Paulette Niewczyk: Well these are the kind of - so when Kathy started our intro and we talked about, you know, subscribers don't pay for use of the measures but they pay for the services, these are the types of customer reports that we provide to our subscribers.

We give them this level of you know, granularity so they can see. And we do this not only by the year, but often by the quarter. So this is what we have been doing. You know, not only do - can it be done, it is being done. If you take a look, you know, just in one year.

So for 2018 there's 914 facilities. I didn't want to, you know, belabor you with all of that information. And then at the same time I also - it isn't clear what exactly it is that NQF is requesting of the measure developer.

So when I first provided some of the information on performance, I did it only for one year. Because it wasn't explicit that you wanted a year-over-year look.

Certainly you know, I was able to do it when requested. But if the signal and noise analysis is what you're looking for, then I think it's just explicitly stating that we would be able to fulfill - I'm speaking grossly in terms of measure developers in general. But we would be able to give that type of information.

Woman: Sharon Cross, do you have any general comments before we move to start?

Sharon Cross: I don't have anything to add. I think everybody has covered it already so, thank you.

Woman: So in that case, are you all comfortable with moving to voting on the first issue? (Terry), yes of course.

((Crosstalk))

(Terry): And this is sort of a strange question but, it has to do with the fact that FIM is intricately part of the payment model. And delta FIM, the change between admission and discharges and important criteria for CMS review. And it comes to…

Woman: Could you move forward to your mic?

(Terry): Oh, I'm sorry. So the question I have is, do you have any data on sort of the admission FIM score divided by quartiles? And compare that to sort of the discharge FIM scale by quartiles by facility.

But what I'm looking for is experience with a strategy for maximizing your FIM change, which is, do your admission FIM score early in the course? You have three days to do it?

So doing it on Day 1 rather than Day 3, do you have data to demonstrate that that's a stable score? Or is there a change? Because what I'm concerned about is there's a potential to confound the outcome measure by the initial input. And I just want to make sure that that's consistent. That's really my question. I'm sorry to go around in circles.

(Ellen Sholts): Paulette, maybe you want to address it from a statistical viewpoint for Dr. (O'Malley)'s question is if there's data on that. And then from a practical viewpoint I would like to make a comment. But Paulette?

Paulette Niewczyk: I'm pretty certain we do have data on that. I do not have it here for you as part of this submission. But I know that, you know, the FIM has been extensively studied so, I'm sure it exists in some capacity.

From a measurement point of view though, the level of functional impairment would be underestimated, not overestimated. So I'm not sure. For instance, if we captured - somebody who's capturing their assessment on, you know, our four that they were admitted so, they're in Day 1, versus somebody capturing it on Day 3, if anything that person on Day 3 is underestimating their functional gain or their improvement or their change score, right.

So there would be potentially, a greater benefit to doing it early on. But it's still the most dependent score.

(Tom): The question is, does the time that the admission FIM score is measured, materially change the difference between the admission and the disparate score?

Kathy Dann: The system captures all assessments. So that's something that we could look at. I don't know that we have it readily available for today's query.

(Ellen Sholts): I think as a practical matter about the three day window for scoring, which I think is interesting, I think every provider setting would try to do the scoring as early as possible, for exactly the reason you're mentioning. You want to - what you're trying to do is describe the patient when you got them. And describe the patient when you've finished your piece of the baton pass in this race towards achieving functional independence.

I think that it would be very difficult. I don't think you'd gain as much as you would lose if you tried to say, this assessment has to take place in the first 12 hours or the first 24 hours. I think the first three days is sort of a recognition of some reality of staffing changes, of patient's health status.

There could be reasons why you might measure admission function a little bit later. But you would never be doing it for gaming purposes. In other words, you wouldn't drag your feet to assess later, to make it look like more gains. Because if anything, it's going to make it look like a smaller gain was achieved.

So I think everybody is looking to make the assessment as early as possible. And that three days is just what's allowed to be practical.

Paulette Niewczyk: It's a great majority are doing it on Day 1. So we do have the information in the data set. We ask the day patients are admitted as well as, the day of the assessment. So we do have that information.

But yes, the large majority of facilities are doing their patient assessments on Day 1, within that 24 hour window. There are some outliers but again, they're likely underestimating that than patient change. They're not gaming or getting any benefits of that as (unintelligible) stated.

Woman: By doing that (unintelligible)?

Paulette Niewczyk: But they're - you know, there absolutely is going to be a little bit of variance there, yes.

Woman: Yes, I have a question. I spoke with one of my colleagues in rehab. And I asked her, given that currently the FIM is used for payment purposes and also for benchmarking, when that goes away as of October 1, that it's no longer the only game in town, so to speak.

So the benchmarking, and it's no longer required for payment, would they continue to use the tool. And the answer was, absolutely not.

So you mentioned early on that you had researched with your client base, and about 50% had said that they would continue to use the FIM instrument. And I was wondering, if you asked them that question directly. How did you ascertain that number?

Kathy Dann: We surveyed - we have 900 and something facilities. We had our Client Services call 300 of them and ask all these questions. And the answer was, as

the text I used for those five reasons that they're going to continue to use it, came right out of that survey.

That non-Medicare payers require it. They use it for joint commission and (unintelligible). That text is - and I summarized them kind of in the order they came in.

That doesn't mean they're going to use it ten years from now. But in the onset, they have to - they feel they need to continue doing some of what they're doing. And our system will allow them to do that. It's not required though.

But about that same percentage was using the - out instrument for quality purposes prior to it being used for payment. So there's a long-term dependence, I guess is the right word.

Woman: Well the discussion that I had was, you know, in terms of costs.

Kathy Dann: Right.

Woman: Just in terms of cost and time and manpower. That if they're already going to be required, and I'm not - trying not to compare, they're already going to be required to complete measures on these same issues elsewhere, that that's the data that they would be using for quality improvement in the future.

Kathy Dann: And some of them answered that.

Woman: Yes, okay. All right, thank you.

Woman: (Peter)?

(Peter): Good question.

Woman: All right, (Brenda)?

(Brenda): Just a quick question.

Woman: Make sure you use your mic closer to you.

(Brenda): Just a quick question. I wanted to know if you have a way of gathering information about variation by patient groups. You know, whether or not there's any variability or if it's, you know, common outcomes.

Kathy Dann: I'm going to let Paulette take that one.

Paulette Niewczyk: Are you speaking in terms of like impairment type. So you know, a patient who had a stroke compared to a patient who may have had a hip replacement? Yes, absolutely.

So we do adjust the data for impairment type. We also adjust the data for severity. Because even among the patients who have had a stroke, all strokes are not exactly the same. Some can be very mild and some could be, you know, very debilitating, certainly deadly.

So yes, we adjust for both impairment, as well as, severity. We use CMG, case mix group, for our severity adjustment. And we also address for age because there's some variability between patients that are in their 40s, compared to those that might be in their 90s.

So the data is, and all of the data that I provided in my measure submissions have been adjusted for that.

(Brenda): My question actually, and I probably should have been a little bit more specific was, focused more on social determinants. And I was trying to see whether or not there are different population subgroups that you know, you're finding that the improvements are greater for whatever reason. I didn't know if you collected that kind of…

((Crosstalk))

Paulette Niewczyk: Yes. So I did look at some of our social demographic variables. In particular I looked at race, sex, and marital status. And there was (unintelligible) differences. And this is just pertaining to our data, the data that I provided here. There were no differences in outcomes. And there were no differences in admission scores.

So I did not further adjust, since there were no differences that existed. It didn't warrant any further adjustment. However, we are limited in terms of how much sociodemographic variables we have in the data set.

In terms of access, geographic access, I'm sure that plays, you know, a very large role. But we don't have that type of information. In terms of, you know payment and you know, physical access, you know, 50 miles away versus 150 miles away, all of those things will you know, likely play a large role in terms of who gets to certain facilities in the first place. We're limited in terms of what we have.

Woman: Thank you. Are we ready to start? Yes? Okay (Jordan), do you want to tell us what we're about to vote on?

(Jordan): Okay. Voting is now open for Measure 2286 for evidence. The options are A, pass or B, no pass.

Man: Two more.

(Jordan): All right, we need one more vote folks. Voting is now closed for Measure 2286 for evidence. The criteria does pass with 13 votes for pass and eight votes for no pass. That's 62% for pass and 38% for no pass.

Woman: Okay, next is 1B, right?

Man: Yes.

Woman: And 1B is gap. Again, are we prepared to proceed to 1B? Okay.

(Jordan): Okay, voting is now open for Measure 2286 for importance performance gap. The options are high, moderate, low, or insufficient. Again, we need one more vote. Okay, voting is now closed for Measure 2286 for performance gap.

The results are high, 2; ten for moderate; three for low, and six for insufficient. This means, consensus not reached.

Woman: So for consensus not reached, we continue through the conversation. Fifty-eight percent, so that's within our consensus not reached gray zone. We will not vote on an overall recommendation for endorsement. And therefore we will not have the competing measures conversation on this pair of measures because we need a decision and the committee has not reached consensus.

So we need a decision before - a recommendation prior to that conversation, so we'll - if it passes everything else, we would have it as a post-comment call.

(Don Casey): Could I make a point of order. The insufficient bothers me because I'm not sure - I don't know actually, if all the members of the committee reviewed the data that was shared in the memo, which did have as I recall, significant data on the performance. Which I don't know, I'm just saying. I'm calling that to question.

The vote is what it is. But if the case is that people haven't seen those data, then I think it would be worthwhile to look at them.

Woman: And that's why we'll discuss at the post-comment call. We will ask you to start again with gap. And then if the measure doesn't reach consensus or doesn't pass something further down - or sorry, if the measure doesn't reach consensus - sorry, we'll ask you to start with gap. And then anything else that may have not reached consensus, we would discuss them as well.

(Don Casey): Yes, my only concern is we're sort of ditching the competing measures discussion, which bothers me. So I'm just…

Woman: We'll have that later.

(Don Casey): Okay.

Woman: So, that will happen.

Woman: I thought you said we would not have the competing measure discussion.

Woman: Not today. We'll have it at the post-comment call. Because we need you to make a recommendation for endorsement. And we can't make that recommendation if you haven't reached consensus.

((Crosstalk))

Woman: We continue with reliability but, we don't make the overall vote.

Woman: Right. We are going - a little confusing. All the way up to the next to last - up to the last vote but, we don't take the last vote. In which case we now move on to.

Man: To reliability.

Woman: Reliability and validity. I thought we'd done - oh, I'm sorry. And the floor is therefore open and (unintelligible) is up.

Woman: Okay, so I'll just give a recap here. I don't know about you guys but, I have a lot of different measure information in my head so, I'm trying to look at notes.

So the Scientific Methods Panel did review this. They ultimately voted to pass this measure on reliability. We have one vote for high, four for moderate, and one for low. I'm just looking at some highlights from the reliability testing. The score level reliability across the facilities was quite high. Intra-cross correlation core vision of .92.

One of the questions that came up from the Methods Panel Review was just asking about part of the methods. Like why is it necessary to do a random sample of 30 facilities, instead of using all 855? And there was a suggestion

for a stronger method to reliability testing as an alternative. So should we do reliability and validity together, or just reliability?

Woman: First reliability and then validity - reliability and then validity.

Woman: Okay. Looking at the comments then from the survey, overall I don't see too many concerns raised about reliability. One question that came up - and I'll say that this one was from me is, that - so you know, it does like that there's an ability to distinguish the best from the worse facilities. But you know, things are tightly grouped in the middle.

And so, you know, I think (Sherry), you brought this up earlier. I also was wondering about the ability to distinguish facilities in the middle. And particularly wondering about like implications for how the information is shared out and acted upon later? So I'll stop there to see who else has (unintelligible).

Woman: I think we've already covered that in the earlier discussion. Are you comfortable with what we have?

Woman: I think so, unless there aren't further questions from others.

Woman: Further comments or questions? Before - (Sherry)?

(Sherry): Yes, I guess - the inter-class correlation coefficient that would give you a sense of reliability isn't done at the between versus in facility variance. And then within patient between items variance.

And so it's done in the split half reliability zone. And so NQF in your guidance, not sure how that's played out now with the mature - you know at

the stage of development of this measure. But hang on before you answer. Because in terms of interpreting the amount of variation that belongs to the facility, the way it's done, you can't really tell.

So in terms of being able to tell differences between facilities, which is where it's being proposed for us, the reliability that we need to understand how that goes, isn't there. Soo you know, how do you guys want to handle that?

(Don Casey): Yes, so I think what you're referencing is the evolution and thinking of the standard where it's going to be. So for the purposes of this submission and the reason that it passed the SMP is because it was being held to the standard as it existed at the time of submission.

(Sherry): Then my question is, why did you use certain facilities? Because if you do spit half reliability, what you wouldn't want to do is limit your ability to kind of get the precision up where you want it. And how those 30 facilities were chosen makes a big difference.

And then so, sort of if you're doing 852 facilities, would you split the sample into two halves at the facility level? Or are you looking at the patient level within facilities, to give us those inter-class correlation decisions?

Kathy Dann: I'm going to let Paulette answer that also.

Paulette Niewczyk: So this is actually what was recommended for us to do from the - from both NQF and the previous committee - Patient and Family Funded Care Committee.

So we had done - and submitted something else previously, before these measures were approved and endorsed. And they recommended that we do this.

So when it came time to submitting for a maintenance review, we thought that's what they wanted last time. We're going to give it to them again. So this was on the previous recommendation.

(Sherry): Can you clarify how you did it? Did you sample - split the sample within…

Paulette Niewczyk: This is a random sample, yes. So it was a completely random sample of our facilities. And this is their average (unintelligible) derived self-care scores. So…

(Sherry): No, I get that. But is it at the patient - did you split the sample within a facility by patients? Or did you split at the facility level, so you're comparing 15 facilities to 15 facilities?

Paulette Niewczyk: These were 30 facilities that were selected at random. And that the patients were compared within those 30 facilities.

Woman: Paulette, did you compare the patients with - did you group them as a facility and then compare each facility against the other 29?

Paulette Niewczyk: Correct, correct.

Woman: Or did you…

Paulette Niewczyk: So both within - so within the facility, within each of the 30, and then between.

Woman: And you could produce the similar data for the over 800 sites right? It would just advised to you to do a sample. Is that…

Paulette Niewczyk: Yes, I mean in 2018 there were 914 facilities. So it's just in terms of the labor and the SPFF processing time. It was cumbersome. So yeas, I mean absolutely we could do this for all of the facilities. But we just selected you know, random sample.

Again, it was at the advisement of the NQF and the previous committee, that a sample would suffice. That we didn't need to do it for the all.

Woman: I'm going to put my (Martin Ness) hat back on. We would like to finish this - with this measure so we can take a break and have lunch. And it will not be a working lunch. Which means you are free to do whatever - forgive me. So we are up to voting on I think, if everybody is comfortable.

(Jordan): Voting is now open for Measure 2286, acceptance of the Scientific Methods Panel's rating on reliability. Voting is now closed for Measure 2286 on reliability.

The committee votes yes for - 20 votes for yes and one vote for no, to acceptance to the rating of reliability.

Woman: Okay, next up, validity. All right, validity. So the Scientific Methods Panel did recommend that the measure pass on validity. There were three votes for high, one for moderate, and two for insufficient.

I don't see a whole lot of explanation of why a couple of the panelists thought that the measure was insufficient on validity. A couple of the points that are

brought up in the summary here from the Methods Panel are things that we had already discussed this morning about correlation of the measure with the FIM instrument from which it comes.

And then correlation to other outcomes such as the choice of community and length of stay. So in looking at the comments from the survey, just skimming through, again it's a lot of things we've already discussed in terms of correlation with health disparities or social determinants of health.

So I think I will turn it over. If there's anything else anyone has questions about validity that we haven't already discussed, that you want to put on the table.

(Sherry): I just want to ask about the endogeneity. Because you're using validity in a sense that you've got a subset of the larger measure and you're associating it with a larger measure including. Did you include or exclude the new - the variables and you're now pulling off to measure what you're measuring, functional sets?

Did you - when you correlated the subset, did you take those out when you made the correlation with the larger FIM?

Kathy Dann: Paulette?

Paulette Niewczyk: When I did the correlation for the self-care measure to the full FIM. So it was the total full FIM summary score. It wasn't each of the individual items of the full FIM. And I was doing this for the criterion referenced, in essence concordance validity.

Is it, if the FIM measuring function and need for help assistance burden of care, so a small subset of items independently of the larger, would there be some consistency?

(Sherry): The problem with that sort of analysis is you've got the thing predicting itself. So what you might want to think about doing is, taking the largest subset - taking the smaller subset out of the other measures and then associating those two instead.

Paulette Niewczyk: We could certainly look at it that way. That is yes, you know, absolutely could be done. But if we're saying that self-care is a component of function, it's not the be all and end all, right. So there's other aspects of function that are not encompassed in the items within the self-care.

So if we do that we'll likely find some consistency. But it won't be very strong consistency. You're going to have to dominate it by cognitive items and some of the more ambulatory items, which is what mobility is capturing.

But we should see you know, a high degree of consistency between just a snapshot, which would be the self-care measure. It's not the only validity measure that was included in the submission. I have predictive validity, construct validity. It was just one measure. Again, this was also recommended by NQF.

Woman: Further discussion. All right. Okay.

(Jordan): Voting is now open for Measure 2286 on reliability. Options are A, high; B, moderate; C, low; and D, insufficient.

Man: There we go.

(Jordan): I apologize. Voting is now open for - well it worked.

((Crosstalk))

(Jordan): So voting is now open for 2286 on acceptance of Scientific Methods Panel rating on validity. Options are A, yes; and B, no. Okay, voting is now closed for Measure 2286 on acceptance of Scientific Methods Panel rating. And there's a unanimous vote for a yes with 21 votes.

Woman: The next area for discussion is feasibility of this measure.

Woman: Okay, so just summarizing the comments from the survey, we have one comment stating that the data question does not (unintelligible) burdensome because the data already exists and is in use.

We have questions about which of the required data elements are not routinely generated and used during care delivery. Since the submission of these data requires a revenue cycle, activities and subsequent billing to CMS the data elements should be ready and available. But are there any data elements which are not available in electronic form?

And then again just reiterating not too many concerns about validity or feasibility rather, I mean I will say we've already had some discussion about the question of sort of what happens after October.

So I propose we not revisit that unless there's any unexplored aspects of that so any further comments from the other discussants?

(Susan Sparcet): Seeing no cards around the table. We'll move on to vote on feasibility.

Woman: Okay voting is now open for Measure 2286 on feasibility. Options are A, high; B; moderate; C, low; and D, insufficient.

Looks like someone has stepped away; move…

(Susan Sparcet): Yes.

Lee Partridge: Yes.

Woman: So one more vote.

(Susan Sparcet): You're going to be - but you're going to be fine.

Woman: Okay, voting is now closed for Measure 2286 on feasibility. The criteria does pass with 3 votes for high, 16 votes for moderate and 1 vote for low.

(Susan Sparcet): Finally moving onto usability.

Woman: Okay so looking at usability at some point I think it was already noted is that at least within this measure information form we're not seeing data on the changing scores over time.

And so for that reason the preliminary rating some (unintelligible) was that the information used is insufficient.

So then looking at other comments from the committee, you know, the measure is - well one statement says the measure is publicly reported and used for accountability. I want to clarify. Is it publicly reported? Like where, you

know, any member of the public for example could look it up on the web site. Is the measure itself publicly reported?

(Susan Sparcet): I'm not clear on that. We give the data to the University of Texas and they (unintelligible) publications. Are you still there Paulette?

Paulette Niewczyk: I am.

(Susan Sparcet): Okay but…

Paulette Niewczyk: I'm not sure what's happening though?

(Susan Sparcet): …most of these are published to our customers.

Woman: Okay.

(Susan Sparcet): So.

Woman: Okay. And in terms of accountability and as I (unintelligible) the CMS use is changing but there are other accountability uses that are continuing on. Is that correct?

Paulette Niewczyk: Well as you mentioned (Kathy) with (Jayco) and curve and…

Woman: Okay. Then so again there were a couple questions that came up about the lack of data on change over time. Excuse me. I think that covers most of the comments that were brought up in this particular region.

I do want to clarify when we talked about the change over time previously was that (Susan) where you clarified the information within the concept memo, could you point us to which page because that…?

(Susan Sparcet): Yes.

Woman: …sometimes that's (really)…?

(Susan Sparcet): They were batches, talk about that. I think probably a number of people didn't realize that it was there. And therefore - well maybe we should - what we suggest is that over lunch anybody who hasn't had a chance to read that data, you may want to take a look at it because it will be relevant and we also talk about the mobility measure this afternoon.

And it might - if to the extent people felt that because they didn't have that in front of them they were uncomfortable voting other than insufficient in the prior vote, that'll be helpful for - we can't go back and revisit that. But it does mean that perhaps there'll be a lower instance of insufficient in the next iteration and when we come to mobility. Okay.

(Ellen Sholts): So on Page 8, Appendix A.

(Susan Sparcet): Yes.

(Ellen Sholts): What is the actual PDF page that will get us there faster?

Woman: Page 8.

(Ellen Sholts): Eight.

(Don Casey): Lee while that's coming up I just wanted to make a point relative to (Ellen Sholts)'s observation that I have in my notes here. And again I didn't cite which document where which is always a problem.

But somewhere I know I wasn't making this up. I noted there appears to be good movement on this measure with the percentage of IRFs with overall scores at or greater than 75%. That has risen from 28% to 39% over the most recent 4 year period. It's in there somewhere. I just don't know where it is.

Woman: It's.

Woman: (Unintelligible).

((Crosstalk))

Lee Partridge: That's the supplemental statement.

Woman: That was…

(Ellen Sholts): Yes.

((Crosstalk))

Woman: …(unintelligible).

(Ellen Sholts): Yes. So that's what's on Page 8. I will note that given the number of measures and then the density of information on each measure having some of the information in assessing these criteria pulled out in separate memo is incredibly cognitively difficult. I mean now that I look it I'm like oh yes, I did look at that.

But, you know, none of us when we filled out the survey had this memo to take it into consideration. So I think some of the going in circles here is just about the cognitive burden on the committee.

Woman: Yes.

Lee Partridge: We.

Woman: We…

Lee Partridge: Right.

Woman: We apologize that since that came in in response to questions.

Lee Partridge: Yes.

Woman: After we had already sent you the PAs sort of, we did what we could. But yes, thank you.

(Susan Sparcet): Sidetracked a bit there but now back to usability and use, our final vote between this and lunch.

Woman: Okay voting is now open for Measure 2286 on use. The options are A, pass; and B, no pass.

Voting is now closed for Measure 2286 on use. The unanimous vote for A, pass for a 21 vote.

(Susan Sparcet): We have one more vote, usability and use, no, just usability.

Lee Partridge: Just use.

(Susan Sparcet): Use, use.

((Crosstalk))

Lee Partridge: (Unintelligible).

(Susan Sparcet): I need my prompt sheet in front of me. Okay.

Woman: Voting is now open for Measure 2286 on usability. The options are A, high; B, moderate; C, low; and D, insufficient.

Voting is now closed for Measure 2286 on usability. And the measure - the criteria does pass with 4 votes for high and 14 votes for moderate, 1 vote for low and 1 vote for insufficient.

(Susan Sparcet): (Peter).

(Peter Thomas): Can I just ask a question please about the - what happens to this now? We're not going to talk about competing measures. But what - we talk about it again on the phone call, do we ask for more data, what's the situation?

(Ellen Sholts): So the competing measures conversation has to be tabled until you've made an overall recommendation on this measure.

So if there is additional information that you want to help you make a decision on this measure let us know. And the developers can provide that

information. We - and we will give it to you for comment, before post-comment call.

And then we will discuss on the post-comment call. The committee will attempt to reach consensus on this measure and to make a recommendation. And then assuming both measures and the pair are recommended they you would have the competing measure conversation.

(Susan Sparcet): Which means that we're postponing that discussion until September, is that correct?

(Ellen Sholts): Yes.

(Susan Sparcet): Please don't throw away your notes.

(Ellen Sholts): We will be providing details, the report and everything. We'll work with our chairs to put together a really solid packet of information for you all on this. And I expect we'll consider scheduling a second call to get through everything because we'll have comments to address as well.

(Don Casey): You do have another pair of measures that we're considering today that will be open for a complete measure discussion. So that's just one pair of measures that we'll have to table for later but the other one is still open.

(Susan Sparcet): First measure up after lunch…

(Don Casey): Yes.

(Susan Sparcet): …which reconvene at quarter to 1:00.

(Ellen Sholts): Take some comments.

(Susan Sparcet): Oh we have to do public comments first and then we will reconvene. All right, open this mike.

(Ellen Sholts): Now open for public comment. If anyone wishes to make a comment, please come to the mike in the room or you can also ask on an open line or make a chat, comment via chat.

Okay no comments have been received so we will go on break for lunch and reconvene at quarter of.

Operator: The conference has been muted.

(Susan Sparcet): If I could ask you all to lift your ear for a minute. We have had a discussion with (Elise) who's left us but she suggests that in view of the confusion about where we could find the data that was relevant for the vote that we - in which we had the massive insufficient evidence response that she would be comfortable with our - if we wish going back and re-voting on that measure.

And if that's the will then this is the way this scenario would play out. We would go back and revisit if then the vote changes so that the measure passes. We would proceed to final vote on the Healthcare Measure 2286. That would then place it back in play for the purposes of the competing discussion.

We have three more measures to go through before we can have that discussion. Namely the mobility measure from UDSMR and the two federal - there are two federal counterparts.

We assume that we would - if we clear all those we would then have the discussion about related and competing and that would end our day. That means that next week we will have all of the CAHPS measures which we haven't touched. And we would also have I think three left over, other measures all - maybe one UDSMR and two CMS.

Woman: (Unintelligible).

(Susan Sparcet): They're all just CMS. Okay. Is there a consensus we should proceed that way?

Personally I think the related and competing measure is the most difficult to do by phone of any I can think of so I would hope we can get there this afternoon.

Okay, then if everybody is back let's be sure first of all you know exactly what was submitted on June 6th that we can look at. Do that (unintelligible) pull the page up. Can we?

Woman: Yes. We can…

Lee Partridge: Oh I know.

Woman: We can do that. Can (unintelligible).

(Susan Sparcet): It's appendix. Yes.

Lee Partridge: So while this is being pulled up, it may be a convenient time for me to speak briefly about the particular criteria that we're going to be re-voting on. So this is performance gap.

And this differs a little bit from the year-over-year type assessment that we do inside of usability. So we spoke about this very briefly.

But just so the committee is certain on what we're considering here, we're looking for the distribution on a single snapshot of the most approximate data that the developer has put in front of us.

And what we're looking for is to make sure that the measure is actually demonstrating a gap between providers by which we're not seeing too much of a narrowing of performance in which case we'd be considering the measure as a potentially topped out measure.

So we're looking for that specifically. And we've already voted on the usability component of it which is the year-over-year component which the committee unanimously passed.

But as we're looking at this which (Jordan) should be pulling up here in just a moment that is the particular information that we are taking interest in.

(Don Casey): Lee as a reviewer can I just make an editorial here that what I saw, my memory was that the interquartile range was substantive and that the delta over time was substantive. So I'm just saying we need to find that data and look at it.

Woman: (Unintelligible).

Man: Here we are.

(Susan Sparcet): Uh-huh. We are for the record re-voting 1B on Measure 2286.

Woman: Voting is now open for Measure 2286 for performance gap. Your options are A, high; B, moderate; C, low; and D, insufficient.

Lee Partridge: Yes.

(Susan Sparcet): I don't have it plugged in yet.

Lee Partridge: (Unintelligible).

(Susan Sparcet): (Unintelligible) because I'm not plugged in.

Lee Partridge: Do I just (unintelligible).

(Susan Sparcet): So I just…

Lee Partridge: (Unintelligible). All right, yes, so here we are.

(Susan Sparcet): Okay.

Woman: Voting is now closed for Measure 2286 for performance gap. The criteria does pass with 18 votes, 1 vote for high and 17 votes for moderate and 2 votes for low.

(Susan Sparcet): All right, in view of that result we now move onto the vote we didn't take on 2286 which is the final recommendation.

Woman: Voting is now open for the overall suitability for endorsement for Measure 2286. The options are A, yes; and B, no.

(Susan Sparcet): And just for the record I should notice that (Sherry) had to leave us so we're now a maximum of 20 people. No.

Woman: Voting is now closed for the overall suitability for endorsement for Measure 2286. And the committee recommends NQF Number 2286 for maintenance for endorsement with 20 votes for yes.

(Susan Sparcet): Okay.

Lee Partridge: Thank you ladies.

(Susan Sparcet): All right, so the next measure up for discussion is number 2321, functional change in mobility score. And our lead discussion is (Linda Murillo). Well (Linda), all yours.

(Linda Murillo): Thank you. So this is a process measure. And the scientific acceptability analysis had originally decided that this was a no pass due to - let me use exact language. It is the missing. Just a moment, I lost it.

(Don Casey): It says outcome on the slide.

(Linda Murillo): Excuse me.

(Don Casey): The slide says outcome measure and I think you - did - you said process or did I misunderstand you?

(Linda Murillo): Oh it - I'm sorry. I'm looking at the form that has actually got both checked so sorry about that.

Okay that evidence has not been submitted. It was the lack of evidence issue. And I have just received that and looked it over at lunchtime. And I'd like to get everybody's input on whether or not they would agree that it's a pass at this point that we have that information. Are we going to…

Man: (Unintelligible).

(Linda Murillo): Has everybody seen the data?

Lee Partridge: This is also data in the memo that you're talking about.

(Linda Murillo): Yes.

Lee Partridge: Yes, okay. So if…

Woman: You can pull that back.

(Linda Murillo): Because…

Lee Partridge: So if we want to pull that back up, that would be fine, just looking for it. Yes.

(Linda Murillo): Yes. I found it to take care of that issue that was the no pass issue.

Lee Partridge: Yes so, just the next page there. Yes.

Woman: Yes (unintelligible).

(Linda Murillo): And so I think…

Lee Partridge: Right.

(Linda Murillo): …it's important that folks see it.

Lee Partridge: Yes. It's really the same issue I think as the last measure.

(Linda Murillo): Yes it is. It's the same issue.

Lee Partridge: So yes. Okay.

(Linda Murillo): Yes.

Lee Partridge: So any other initial thoughts?

(Linda Murillo): Well I love that.

Lee Partridge: Well.

(Linda Murillo): Yes so…

Lee Partridge: Things that would affect our voting.

(Linda Murillo): Sure. So there were some concerns noted. There was a narrow, well I'll table that for later.

Lee Partridge: Narrow path.

(Linda Murillo): Yes. So we're going to just vote now on whether or not there was sufficient data for the (pass). Okay, anybody else have any…?

Man: (Unintelligible).

(Linda Murillo): …comments?

Woman: Voting is now open for Measure 2321 for evidence. Your options are A, for pass; and B, no pass (unintelligible).

(Susan Sparcet): We do. Actually we - (Brenda), I mean (Linda) has three colleagues who were also working, looked at this measure. And before, excuse me, we proceed to voting on 1B (Brenda), (Terry), (Linda), I mean (Brian)…

(Brian Limburg): Yes.

(Susan Sparcet): …any comments you'd like to share in general about this measure?

(Brian Limburg): Yes, thank you. And it is really a general comment in the sense that I think this kind of information is very important to consumers.

But there's a history behind consumers going into facilities and needing to be able to - the facility needing to be able to show that there's improvement for the therapy to continue.

So I just wanted to mention that. There's a case in Medicare that (unintelligible) which basically said that the individual if their status, their health status is maintained it doesn't have to necessarily improve for coverage.

So when I was - when I went through this that occurred to me that one of the things you don't want to do is only measure when there's improvement, that you need to be able to have measures that show that either the decline has been reduced potentially or that the individual has stabilized because of the

therapies so very general comment but I think that from a consumer point of view these kinds of measures are important.

(Susan Sparcet): (Terry) no comments. (Brenda) any general comments on this measure? No.

All right, and are we ready to proceed to 1B?

Woman: Before we vote for 1B I just wanted to announce the results for evidence for Measure 2321. There was 18 votes for a pass and 2 votes for a no pass.

(Susan Sparcet): All right.

Woman: And with that I open voting for Measure 2321 for performance gaps. Your options are A, high; B, moderate; C, low; and D, insufficient.

Lee Partridge: (Unintelligible).

Woman: Okay voting is now closed for Measure 2321 on performance gaps. The criteria does pass with 1 vote for high, 17 votes for moderate, and 2 votes for low.

(Susan Sparcet): All right. I forget. Do we have - we would move onto evidence.

Lee Partridge: (Unintelligible).

(Susan Sparcet): Yes and I'm sorry.

(Ellen Sholts): Reliability.

Lee Partridge: Yes, the (unintelligible) reliability, reliability.

(Susan Sparcet): My voice keeps leaving.

Lee Partridge: Okay.

(Susan Sparcet): All right.

(Ellen Sholts): Okay so the panel found that reliability, there was one high, four moderate, one low. And my concern prior to reviewing the data was or I'm sorry, prior to your answer to the questions about self-care was why 30 instead of running a much larger sample size, but you've already answered that question.

So really all of my questions have been addressed after speaking about the self-care and I'm wondering if anybody else on my group has any comments on reliability.

(Don Casey): I have a question. Did you and your colleagues have sort of the same concerns that I noted in the other measure regarding the (unintelligible) of ways to address social determinants in health and disparities. Did that - did you…?

(Ellen Sholts): Yes.

(Don Casey): Yes, okay.

(Ellen Sholts): Yes.

(Don Casey): Because it seems like this would be one of the areas that would be substantively impacted by.

(Ellen Sholts): Absolutely.

(Don Casey): Yes.

(Ellen Sholts): And that was - actually thank you. That was one of the things that I did have concerns about that was the - I believe the population that you captured was very narrow. And I was - in terms of socioeconomic status and all of the other social determinants of health.

And so I was wondering if you could provide some additional information on that or what the reasoning was.

(Susan Sparcet): Paulette.

(Ellen Sholts): Excuse me.

(Susan Sparcet): Paulette is going to…

(Ellen Sholts): Oh.

(Susan Sparcet): The one that couldn't get her - on the plane (unintelligible).

(Ellen Sholts): Okay.

Operator: The conference has been unmuted.

Paulette Niewczyk: I think I'm muted now. Yes. Okay.

(Susan Sparcet): You're unmuted.

Paulette Niewczyk: We're - yes, we're limited by what's available in the data set so we do have access to race, certainly sex, we have age, we have marital status, we have payer and sometimes payer source could be a proxy for affluence or income means, financial means.

So I did look at the data for - it's already age adjusted. But I did look at the data for sex, race and marital status. Marital status is a proxy for social support because that would certainly potentially affect discharge to community if there isn't anybody, you know, in place to provide care that could be needed.

And there was no difference. So there was no significant difference in outcomes on…

(Susan Sparcet): Could you talk up for us please? Yes.

Paulette Niewczyk: Pardon.

(Susan Sparcet): Could you speak up?

Paulette Niewczyk: Oh I'm sorry. Yes. So when I looked and at the outcomes based on race, based on marital status and based on sex there were no differences so it was not adjusted for those variables.

Now we don't have income. We don't have, you know, education, highest level completed. Those would all be excellent variables, you know, to look at. And I'm just certain that they would have some level of influence. But we just don't have it available in the data sets.

(Ellen Sholts): Okay. One of my other concerns had been the lack of change to any of the categories that you're capturing over the years. It's been very consistent. But it's also not changed.

And I was wondering if you had planned any updates that might be able to capture more of that information.

Paulette Niewczyk: What sort of change were you talking about? I'm not sure.

(Ellen Sholts): Just looking into more of socioeconomic.

Paulette Niewczyk: Like from the change over time the mobility measure like the facility distribution, there has been change there…

(Ellen Sholts): No.

Paulette Niewczyk: …many more (unintelligible).

((Crosstalk))

(Susan Sparcet): To get at more data.

Paulette Niewczyk: In 2015.

(Ellen Sholts): Excuse me.

(Susan Sparcet): I mean.

Paulette Niewczyk: Only 16.7% were in the top performing, 75th quartile or 3rd quartile. And in 2018 we had 24% so we are seeing some change at the facility level so performance is improving.

Woman: The upper end of the range again is going to end when you - when a patient is good enough to be discharged so.

Paulette Niewczyk: Right.

Woman: That's going to affect one end. So I think the change we're looking for is to see more hospitals getting more patients into that higher change range.

But they can't extend the overall change for any one patient because it would be improper to keep a patient longer than they need to be in a rehab hospital, you know, to achieve some higher scores like be able to walk 150 feet instead of, you know, some shorter distance, right.

And to your other point about looking at the socioeconomic impact, I think the key thing is to have a measure that you trust that measures change of function which we believe we've shown evidence that this is a very trustworthy measure.

And then as you get other data elements and proxies for example everybody is struggling with this throughout all of measurement, are home zip codes for example a good proxy for economic status or whatever.

But if you have the measure that you like and it's good and strong and then as we find more pieces of data to marry that with to then say is a hospital who serves a lower economic population performing less well.

And then is that because it's a lower economic population or some other reason, you know, all of those other…

Paulette Niewczyk: Right.

Woman: …things. But fundamentally you need a strong measure and then marry it with socioeconomic social determinants.

(Ellen Sholts): Okay thank you. I don't have anything else. Yes.

(Terry): I'm (Terry).

Lee Partridge: Could I ask a question and I'm not exactly sure if it goes in this section. But I reviewed…

Woman: Yes.

Lee Partridge: …a similar measure but I'm going to be the principal spokesperson I guess on a similar measure but it involves changing functional status compared to a risk adjusted mean at discharge.

And I'm just trying to get a sense for what the relative value is of each one of these measures. One that you measure functional status between admission and discharge and one where you're measuring functional status at discharge compared to a risk adjusted mean.

Are we getting at the same thing or are they really geared toward different outcomes or data?

(Susan Sparcet): Paulette.

Paulette Niewczyk: Our measure is meant to capture that individual patient's change from admission to discharge with the theoretical understanding that that level of care that they're receiving during that period of time is impacting that change. It's - if it's improvement in function then the services that they received led to that improvement in function. Sometimes patients don't change, sometimes they decline and perhaps there's another reason. Maybe there was a fall or there was an infection.

But it's measured to capture that change. What it sounds like to me from the other measure that you're referring to is it was compared to a mean, meaning an expected value of where the patient should have been. This isn't doing that. Our measure isn't doing that. Because I would have to say well what is the expectation. It's going to be different based on age. It's going to be different based on what the impairment was. It was - it would be different based on what precipitated the rehab stay, comorbidities. So it just takes each individual as their own. It's not norm referenced for instance.

Lee Partridge: Thank you.

(Ellen Sholts): It creates a database does it not Paulette in which you can take that next step. In other words you could risk adjust and say of all the patients who had a certain diagnosis and a certain risk profile, how many of them gained, you know, how many points in functional measure.

And then again something else that can drive clinical performance of, you know, the average person at the average hospital would gain this much. Let's see if we can't do better.

Is - but the measure itself as it's, you know, before you does not do that. But it's a stepping stone to drive that kind of change.

Paulette Niewczyk: Correct.

(Susan Sparcet): Does that answer your question?

(Don Casey): It does. I'll have some follow-ups when I talk about the measure that I've been assigned.

(Susan Sparcet): Good, thank you, any further discussion? If not, (Brenda).

(Brenda Lee): Well I just wanted to say I was going to ask (unintelligible).

(Susan Sparcet): Microphone.

(Brenda Lee): Oh. Pardon me. I was going to ask a question about the variations by health condition or impairment. But you started talking about it so I didn't need to do that.

But I think that that kind of information is helpful from the standpoint of if you're trying to employ a population health kind of intervention in the settings that you're working in then I think that that would be a way of, you know, at least getting some concrete data to be able to use this as a benchmark for the different comparisons that could be made.

Paulette Niewczyk: Absolutely. Yes, definitely. And we do do that with our subscribers so they do have that standard reference that they can use for things like goal setting during that therapeutic process. But that's just not a component of the measure as it sits in front of you. Yes.

Lee Partridge: Great. So I believe we're at the point of reliability and validity. See if we want to accept the methods panel of moderate for one and high for the other I believe. Getting four, two more, yes, there we are. Well let me turn one on.

Woman: There we go. Voting is now closed for Measure 2321 on accepting the Scientific Methods Panel's rating for reliability, 20 votes for yes and 1 vote for no.

Lee Partridge: Okay and any other discussion of validity? I think we've been through that, right. Okay, great so that's done.

Woman: Voting is now open for Measure 2321 on accepting the Scientific Methods Panel rating for validity. Your options are yes and no.

Okay voting is now closed for Measure 2321 on accepting the SMP's rating for validity with yes, 18 votes for yes and 3 votes for no.

Lee Partridge: So feasibility.

(Ellen Sholts): Yes. So under feasibility the CMS portion, they're no longer going to be using the FIM and they're changing their measures.

And that was one of the concerns that I had had about usability. But you've explained it, you - with your survey that there will be still an important function for this tool.

So I would like to get the input of the other members of my team. But I had no concerns beyond those. Thank you.

(Brenda Lee): I did not have any concerns because of their sales track record and using the instrument about the feasibility of it. So I don't have any further comments about it.

Lee Partridge: Yes, thanks.

(Ellen Sholts): I did want to clarify so the information you shared previously about sort of the percent of your subscribers who indicated they would likely continue using the FIM that was the entire FIM instrument, right. So it would include the items pertinent to this measure as well as Measure 2286.

(Brenda Lee): They didn't always distinguish. It could be part of the FIMs. Some of the issues was the cognitive. They didn't have anything else for cognitive. So when I made my statement I said FIM or portions of it. So we don't know that distinction.

(Ellen Sholts): Okay.

(Brenda Lee): Exactly.

(Ellen Sholts): Okay.

Lee Partridge: Okay any other comments? Okay feasibility.

Woman: Voting is now open for Measure 2321 on feasibility. Your options are A, high; B, moderate; C, low; and D, insufficient.

Voting is now closed for Measure 2321 on feasibility. The criteria does pass with both 3 votes for high, 16 votes for moderate, and 2 votes for low.

Lee Partridge: Okay usability and use. I know we've discussed a lot of things that have to do with the other measure in this group, anything different.

Okay should we go vote on that?

Woman: Voting is now open for Measure 2321 on use. Your options are A, for pass; and B, for no pass.

Voting is now closed for Measure 2321 on use with a unanimous vote for pass for 21 votes.

Lee Partridge: Okay any other comments about usability?

Okay, let's vote on usability or (Linda) did you have - no, okay. Let's vote.

Woman: Voting is now open for Measure 2321 for usability. Your options are A, high; B, moderate; C, low; and D, insufficient.

Lee Partridge: Okay. We may be there since (Peter) had to leave for a little bit.

Woman: Voting is now closed for Measure 2321 for usability. The criterion does pass with 1 vote for high, 16 votes for moderate, and 3 votes for low.

Lee Partridge: All right, and overall suitability.

Woman: Voting is now open for Measure 2321 on the overall suitability for endorsement. Your options are A, yes; and B, no.

Okay, voting is now closed for Measure 2321 for overall suitability for endorsement. The committee does recommend Measure Number 2321 for maintenance of endorsement with 20 votes for yes.

Lee Partridge: All right, shall I just keep on going with the CMS, some more measures to this then?

(Susan Sparcet): Yes.

Lee Partridge: Okay thanks.

(Susan Sparcet): Thank you.

Lee Partridge: Great, thank you.

(Susan Sparcet): We're sorry (unintelligible).

Woman: (Unintelligible) fun, no way.

Lee Partridge: Fun, it's a different kind of fun. All right so it's 26. 2633 is the self-care. Yes.

(Susan Sparcet): (Unintelligible) 2634.

(Ellen Sholts): Are we going to do 2632 first or…?

(Susan Sparcet): No.

(Ellen Sholts): No.

Lee Partridge: No.

(Ellen Sholts): Oh okay.

(Don Casey): We're not going to do (that).

Lee Partridge: Yes, we have to do the potentially…

(Ellen Sholts): Okay.

Lee Partridge: …competing measure…

(Ellen Sholts): Okay.

Lee Partridge: …first just…

(Don Casey): And just those two and then (unintelligible).

(Ellen Sholts): Okay.

(Susan Sparcet): (Unintelligible).

(Don Casey): So just a point of clarification, right. Sorry if it wasn't clear. So we're going to be doing the - we have a total of five CMS measures that were submitted that are all fairly similar. However we're only going to review the first two that are related and competing with UDSMR's measures.

And then we'll move into our related and competing discussion and our hope is to be able to complete that by the end of the day.

And whatever is remaining, i.e. the CAHPS measures and the remainder of CMS's measures we will complete telephonically at a later.

Woman: On there.

(Susan Sparcet): No, 2635, 2635 and 2636 are related measures but they are not competing measures.

And welcome back to our team. You spent a lot of time with us a few years back. I think we probably asked you more and more questions.

But do you want to follow the same process we did with the other two?

Man: Okay.

(Susan Sparcet): And do both measures…

Lee Partridge: Those measures together.

(Susan Sparcet): …at once.

Lee Partridge: Okay.

(Susan Sparcet): Yes, so go ahead.

(Ann Beach): Okay great. So first off, thank you for the time, that this panel and the committee for reviewing all the materials. We looked through the worksheets and as part of my intro I'll be happy to address some of the questions that came up. I know obviously you reviewed things carefully. You had some great questions.

So my name is (Ann Beach). I work at (Archway) International. I'm a nurse by training and I have a Ph.D. in epidemiology.

And I'll let Alan do a quick intro.

Alan Levitt: Yes. Hi. Alan for those of you who don't know me, Alan Levitt and I'm the Medical Officer in the Division of Chronic and Post-Acute Care at CMS.

(Ann Beach): Great thanks. And there's also a team of people on the phone. I won't go through everybody. But we have an Interdisciplinary Team that works on this, all of these measures.

So to begin with the Measure 2633, 2634, looked at change in self-care and change in mobility again focused on patients receiving inpatient rehabilitation care. These measures have been implemented as the IMPACT Act measures so for those of you not familiar that's basically the law that Congress directed the Department of Health and Human Services to implement quality measures in the area of change in self-care and change in mobility.

So there's 7 self-care activities and 15 mobility activities. In terms of the area of importance we described in our documentation therapy interventions as a key way that inpatient rehab facilities can improve their outcomes. We also showed some gaps in care in terms of differences in outcomes related to race, ethnicity at the patient level in the literature as well as some facility level differences for example differences in different parts of the country in terms of outcomes.

In terms of reliability, validity we provided I think a lot of analysis. We focus on some analyses related to data elements. And then because this is an

instrument-based measure, well these are both instrument-based measures we provided also scale level analysis which is really the observed self-care scores, observed mobility scores either discharge or change.

And then we also provided some analyses at the QM score level so that's the risk adjusted data.

In terms of feasibility as I believe you're aware these measures have been implemented in all IRFs in the U.S. and that was actually data collection starting in 2016 so that's obviously a big update since the last time we were at NQF.

In terms of usability and use there are confidential feedback reports that all of the inpatient rehab facilities get at this point in time. And CMS has finalized public reporting of these measures in 2020, which is next year based on Calendar Year 2019 data so that basically is the data being collected now. And that has been finalized through regulation.

In terms of interoperability all the items, actually in all of our measure, all of the CMS measures have link codes so they are interoperable from that standpoint.

In terms of improvement, I guess one of the things I'd like to mention because I know there were questions about this, in the last few years because of the IMPACT Act there has been a lot of things going on within inpatient rehab facilities.

And so they have been focused on issues related to additional payment reform, things about other alternative payment models.

And so it's important for measures like these to be part of the healthcare delivery system because as additional payment reforms happen in the future there would be concern at least in my mind that as payment models perhaps try and reduce costs and facilities respond to that there may be shorter lengths of stays and patient functional outcomes might be impacted negatively.

And so even if there's not improvement I do think that there is a role for having functional improvement measure in order to demonstrate that there's maintenance of function and that quality doesn't decrease.

And there is historically, actually this is part of our original application. We did demonstrate that for patients with stroke there was a study where somebody or (Brian), our researcher in Rochester, showed that as length of stay decreased in IRFs over time functional improvement also decreased. So again an important thing to have measures focused on functional improvement.

Let's see. Let's see. The other thing I wanted to mention is that we have made feedback based on - we've made changes to our measures based on feedback from the industry as well as other stakeholders. And that we have also reviewed all the specifications since our last original endorsement. We've reviewed all the other existing NQF endorsed measures to see where we can harmonize.

And so we tried to summarize some of that. So for example we changed to a different approach when we do the risk adjustment as opposed to a racial approach when we looked at the data. We also did that because we have analogous measures or aligned measures in the skilled nursing facility setting. And there's a wider range of functioning in facility level scores that you

would expect. And so we wanted to align with what made sense for them for that setting.

In terms of social risk factors, when we put together our risk adjustment model, we considered whether there was a conceptual or theoretical model. So for example you'll see that we do adjust for age in all of our risk adjustment models because there's a biological basis for that. There's a lot of research obviously. People who get older have less kind of reserve basically in terms of being able to improve function and our analyses obviously show that.

For other factors for example race, ethnicity there's not really a theoretical basis why there might be differences in functional outcomes. And if we adjusted for race, ethnicity for example, Asian or black race, that would mean that the outcomes for those patients would be - that we would expect less functional improvement for those individuals. And we - CMS I think has a position that everybody should have good outcomes. And so that's why we don't adjust for that.

In terms of SES, we actually found in our - in these measures that people with lower SES had better outcomes. And so again we decided not to adjust for that.

One question that came up on these measures is related to Medicare being the population. And with interest in understanding why all payer wasn't being used.

So CMS obviously can't just decide to collect data and just implement it. So in fact they have been interested in having all payer data and have actually proposed that. It's currently under rulemaking so that's how CMS would be

able to have all payer data and so that's currently a proposal and in rulemaking.

Let's see. There was very little missing data and our data is less than .01%. For the IRF measures I think there was maybe a bit of confusion. Somebody made a comment in one of the documents saying there was some missing data but it was very, very little.

Also there was a couple comments about the results from the Technical Expert Panel. And so there was a question about why it wasn't higher.

So when we pulled together expert panels, CMS has asked our TI to make sure we hear from all stakeholders. And so we actually invited people from our Competing Measures Group and also people who have actually been provided feedback in the past.

And so we actually did want people who were going to give us feedback about the measures. And so that's why that might be lower.

In terms of exclusion criteria there were some questions about the exclusion criteria for people with incomplete stays. So basically if a patient is in an IRF and they have a medical emergency they will be discharged quite urgently back to acute care.

And so when we were working on developing the items and developing, you know, the measures we were concerned about basically, you know, somebody needing to get emergency care. And so you can't really do a functional assessment at that point in time.

And we asked providers in multiple settings including IRF what do you code if somebody has a medical emergency? If you have to put a code what do you put?

And some people will report what was actually the person's status prior to the current medical event and some people were coding low and so it wasn't really consistently done. And so the feedback was that basically not collecting the data was better than not having reliable or consistent data.

Lee Partridge: Can you please start wrapping it up? That would be great, thanks.

(Ann Beach): I'm sorry.

Lee Partridge: We're just - if you could pick (it up) a little bit.

(Ann Beach): Yes.

Lee Partridge: Thank you. Yes.

(Ann Beach): I think that's it.

Lee Partridge: Okay.

(Ann Beach): So I think Alan, is there anything else?

Alan Levitt: No.

(Ann Beach): Okay thank you.

Lee Partridge: Thank you. Okay. And I believe the discussants for this one is (Brenda) and then (Peter) when he gets back and then (Sharon) and (Don).

(Brenda Lee): Thank you for that very comprehensive overview. As I looked at this measure I thought of myself as a patient that might need it. And having access to information that shows, you know, what the performance is across facilities is very important. Because it's often the case that patients and families are navigating the score without some prior - the benefit of prior experience.

So I thought that this was an important measure. I have a question and, you know, you mentioned just a few moments ago about the outcomes of patients who are from lower socioeconomic backgrounds. Could you elaborate just a little bit on why that is, well what the thinking is around those kinds of outcomes because that just seems unusual?

(Ann Beach): So let's see. So we did use the ARC SES Index and had five groups basically. SES Higher Index was associated with higher SES. And we - in our results we did find that patients with lower SES had better outcomes relative to the higher SES.

So some of that could - I'm not sure that we really know. But possibly IRF patient rehab facilities do have an adjustment for - to (DISH), but based on number of the Medicare I think.

But there is an adjustment for payment related to low income. And so it's possible that facilities who, you know take care of those patients, you know, are reimbursed basically to be sure that patients are really ready to go home.

I do think - I think somebody else actually mentioned that. Yes, they could have longer length of stay and therefore have better outcomes. I think

somebody else noted that living alone; patients living alone had better outcomes. And it is true. I can speak to this one that patients who live alone IRF will keep them a little bit longer to be sure they're really ready to go home.

So if somebody is at risk maybe for a readmission or whatever or really not being ready to come home, I think IRFs are very person-centered when they provide care and would make judgment to do that.

And Alan is a physician. He used to work at a rehab facility can…

Alan Levitt: Yes. I mean I think that as you know there are many factors in terms of when a patient can be discharged from a post-acute care facility. And certainly if somebody is being discharged to a lower SES community maybe it takes them more days to mobilize the resources necessary for that discharge.

And so this could just be an artifact of longer length of stay. It's something we follow. I mean again this is the reason we don't just routinely adjust for SES when we find it because we see things going in all different directions.

And so that's why we really need to continue to figure out what we should do whether we should be risk adjusting, stratifying, watching, I mean all these different things.

(Brenda Lee): Thank you for that.

Lee Partridge: I guess we got a - (Ellen Sholts).

(Ellen Sholts): I was just going to add that I think I've seen this with patient reported outcomes as well. So when you're in a vulnerable population that's actually

starting off lower and you're looking at change, if you're starting off worse that group can have larger change scores. So you may as Alan mentioned, be seeing a fair amount of people who started much worse than the comparison population.

Lee Partridge: (Deb) you had a comment and then (Don) you're one of the discussants so I'll let you go after that.

(Deb): I had actually wondered if it was an access issue. Because I wondered if perhaps only selective patients were getting referred to these rehab facilities.

But I looked at the data that was in my measure which is 2635 which was a different measure but it's the same data. And the percent of non-whites is very similar to the general population. So I don't think it's a difference in the referral pattern at least at this - at a global level.

Lee Partridge: (Don) did…?

(Don Casey): No. I'll pass.

Lee Partridge: No. Okay. We should just let (Sharon) and (Don) weigh-in on any other comments because they're the discussants.

(Sharon): Don't have anything in addition to that.

Lee Partridge: Okay (Beth).

(Beth): Yes. Just had a question on why there were - and maybe get into discussion with the competing measure. But there were different measures. What the reason was for developing new measures and then also this had change in self-

care with the cognitive function, not being there. And if that's a more appropriate question for later, I'm happy to bring it up at a later point.

(Ellen Sholts): I think it does get to the importance of the measure report. I'd like to hear (unintelligible).

((Crosstalk))

Lee Partridge: Okay. Okay.

Alan Levitt: And again. Yes.

Lee Partridge: Okay.

Alan Levitt: I'm sure we'll talk about this more later. But the change in self-care, mobility measures were measures that were mandated through the IMPACT Act that they need or the domain that was there. And again the items that are part of our measure are the standardized assessment items. Again that also mandated by the IMPACT Act.

Lee Partridge: Okay, good, any other questions about evidence? (Ellen Sholts).

(Ellen Sholts): So I just know that in the evidence summary it's noted that most of the IRF research on functional outcomes focus on motor function which encompasses both self-care and mobility.

And so again here I feel like this is a place for why are we splitting things out when we could be combining together. I recognize that, you know, it may reflect some of your mandate and certainly applies to both these measures and the others.

But I do think it's worth pointing out that when the evidence-based shows that they should be kept together, why are pulling them apart.

(Ann Beach): Yes. So I'm happy to address that. And that's a great question. So within inpatient rehab facility settings there's a wide range of patients, some people who have experienced a neurologic condition for example, a stroke, and other people who have experienced an orthopedic condition.

And so when we actually have done some of the analyses early on actually we found that when you merge all the data together across these very different diagnosis groups you actually have combined improvement that kind of de-mapped a little bit overall. When you split into self-care and mobility you actually have two different sets of constructs basically.

And I think, you know, most - I would say, you know, in the inpatient rehab facility world originally the idea was that you could measure motor function and cognitive function together.

And then the research showed that across diagnosis groups it's actually not a good idea to put it all together because improving a cognitive function and improving a motor function are very different.

And then when we started looking at the data elements when we had this data from the demonstration project we found that within diagnose - across diagnosis groups splitting self-care and mobility meant that you were measuring those two constructs more precisely than if you merged them together.

And you may have patients who have a lot of improvement that they might be getting in the mobility area and not in the self-care area. And, you know, one of the reasons cognitive is actually a separate construct is because you might have individuals who don't have cognitive issues. And so they're not going to show improvement.

And so that's why we wouldn't merge them together. So I hope that helps.

Lee Partridge: Okay great. Are we ready to vote on evidence for this?

Woman: (Unintelligible).

Lee Partridge: Oh.

Woman: (Don).

Lee Partridge: Oh sorry. I'm sorry (unintelligible) okay.

(Don Casey): So couple things. I always get confused when you guys - you measure developers use the term socioeconomic status. Because I think you're relying a lot on administrative data.

Can you - I think a related question is sociodemographic differences. I don't know if that is a real term that you think about or use. But can you talk about that factor in terms of your analysis?

And let me reflect on something that I think is appropriate here. In the memo it was noted in the 2015 Standing Committee Voting Summary which is on Page 65, the committee inquired, that the committee then inquired about the lack of information on disparities. You measure performance and the

developer, you, I think indicated that data is available. However due to the wealth of information, you weren't sure about how much and what data to submit. You agreed to provide additional information specifically on age, race, payer source, during the public comment period.

But I didn't know of or see where or when this additional information was provided back then or if it was summarized in your report. So that's kind of maybe a longwinded answer of saying I'm glad you answered the question about SES.

But I'm still a little bit unclear about it. I don't think we need to get into it.

But I do want to address the sociodemographic question.

(Ann Beach): Sure. So I'm not sure that actually that was our measure that…

(Don Casey): Okay.

(Ann Beach): …you were referring to but…

((Crosstalk))

(Don Casey): (Unintelligible).

(Ann Beach): …either way happy to answer your question. So we did see differences in terms of race. We didn't see differences in ethnicity. But we did see differences in race, individuals who were black, African American as well as individuals who were Asian had lower functional improvement both self-care and mobility relative to white. So is that one of the questions?

(Don Casey): Well and what about demographics by sort of location as site, you know, zip code, MSA, whatever?

(Ann Beach): Okay. So the ARC SES Index…

(Don Casey): Rural.

(Ann Beach): …is based on zip code so that would've been covered in our ARC SES analysis. And that again showed lower SES actually had better outcomes.

(Don Casey): By zip code.

(Ann Beach): Originally based on zip code, yes.

(Don Casey): And in rural areas no different than urban areas.

(Ann Beach): I'm going to actually ask my team. So (Molly), (Lauren), did we look at rural?

Woman: (Unintelligible) but (unintelligible).

(Ann Beach): Oh actually we did. Yes, sorry.

Man: Yes.

Woman: (Unintelligible).

Man: Yes. Yes.

(Ann Beach): Yes. Yes. It's really, yes, not significantly different. Sorry. We have the cognitive problem too that we have so much information. I'm trying to remember what we did.

So yes, so we looked at race, ethnicity, living alone, IRF in a city and the ARC SES based on zip code.

(Don Casey): Yes. And then quickly the data on the performance for two years, there was one table where you had decile rankings but then you summarized it. And I just want to be sure I am clear on it. The two year data that you summarized as I recall reading didn't change much between the two years you had presented it.

(Ann Beach): That's correct.

(Don Casey): Okay thank you.

Lee Partridge: Okay anyone else I missed?

Okay, should we vote on evidence for 2633?

Woman: Voting is now open for Measure 2633 for evidence. The options are A, pass; and B, no pass.

Lee Partridge: Okay so we're good then. Okay.

Woman: Voting is now closed for Measures 2633 for evidence. The criterion does pass with 17 votes for pass and 2 votes for no pass.

Lee Partridge: Okay, the discussion of gap for 2633. And our other discussants, any comments on gap? Could you use your mike please (Brenda)?

(Brenda Lee): The developer did provide information about the performance gap. And, you know, I did not have any concerns there. I don't know what the other reviewers you want to say about that. I sense we thought there was opportunity for further improvement.

(Don Casey): Great. This is Medicare only, right. So you don't have performance data for non-Medicare in your model, right? So in that sense there's not a gap as much as there is a blank spot if I'm on Medicare.

(Brenda Lee): I guess what I was looking at and maybe and my focus might have been too narrow. I was looking at the difference in the population group. Because if I'm not mistaken there was a linguistic difference between non-English speaking African Americans and African Americans in other groups that I'm not sure about, that I don't recall.

So just in that little example I felt like there was opportunity to actually (make) some improvement there. And that's why I asked the earlier question about the SES, you know. I made some assumptions perhaps. But I thought there's opportunity (unintelligible).

Lee Partridge: Okay, any comments on performance gap? (Ellen Sholts).

(Ellen Sholts): Can I just ask a question? Is there like a maximum theoretical for like just to understand the possible range? Like you can see the actual range you observed. But, you know, where does that fit within what would theoretically be possible?

(Ann Beach): Yes. So for the self-care measure it could theoretically range from minus 35 to positive 35. For the change in mobility it could be negative 75 to positive 35.

And the average, what's for self-care, is something like 11.2%, 11.8% and that change in mobility was about 27 units.

Lee Partridge: Great, anything else about performance gap? I think we pretty much covered what was written.

Great, let's vote on performance gap.

Woman: Voting is now open for Measure 2633 for performance gap. The options are A, high; B, moderate; C, low; and D, insufficient.

Voting is now closed for Measure 2633 for performance gap. The criterion does pass with 15 votes for moderate and 4 votes for low.

Lee Partridge: All right, so now we've got reliability and validity. And (unintelligible), all right so the Scientific Methods Panel reliability was high 4, medium 2. Validity was high 2, medium 3, low 1, so anything (Brenda) that you noticed during the other discussants?

(Brenda Lee): Just that the developers used standardized data. So it would seem like it would be, you know, like the measure. (Jay) would replicate this. And it has - it reflects in my opinion the ability to measure what it's (unintelligible) measure.

Lee Partridge: Others. We're okay with the second metrics. Okay. All right, should we vote on accepting the Scientific Methods Panel? Do that, okay.

Woman: Voting is now open for Measure 2633 for accepting the Scientific Methods Panel's rating for reliability. Your options are yes and no.

Voting is now closed.

Lee Partridge: Okay.

Woman: Yes. Voting is now closed for Measure 2633 for accepting the Scientific Methods Panel's rating for reliability. It's a unanimous vote for yes and 19 votes.

Lee Partridge: Great. Yes, let's continue, any discussion of validity before we vote on that? (Ellen Sholts).

(Ellen Sholts): Oh I have a question about one of the exclusions. So Exclusion Number 2 is patients who are independent with also (unintelligible) disease at the time of admission. So the argument is that, you know, patients already have the highest score so there's not really room for improvement. However there's room for decline.

So if they, you know if they score like independent on self-care (assuming) they are not admitted to the IRF for self-care. So there's something else that's the area of focus. If you're focusing lots of attention let's say on cognitive functioning or motor functioning, you know, could there either be sort of side effects, like unintended consequences that leads to a decline in self-care or it is not attention to self-care where there could be a decline over time.

So I don't know if you explored that in the data. I'm just thinking about it theoretically.

But, you know, we don't want functioning to slip, right, particularly if there's attention somewhere else. You certainly want to be able to maintain independence on self-care.

(Ann Beach): Yes, great question. So first of all, I'd like to mention that those patients are included in 2635, 2636 so they are in the other measure. And basically the rationale is I think you were hoping you did understand the rationale that mathematically somebody cannot actually improve function if they're already at the top of the scale. There's very few patients who are actually at the top of the scale, self-care and mobility.

And basically if somebody is walking up and down 12 steps independently for example for mobility, I'm not sure they'd meet the criteria for being needing to be admitted to an inpatient rehab facility.

You know CMS actually under their coverage requirements require that people have functional improvement as a goal basically. And so we see very little data basically where people are right at the top of those particular items.

(Ellen Sholts): But so my point is that someone could be admitted to the IRF. They could have deficit in mobility for example. But they score independent on self-care.

So when it comes to looking at the quality of care by the IRF like are we going to basically let them off the hook if the self-care functioning declines over time?

So what I'm saying is like should you be looking at the change and score over time that it's either maintained or improved rather than looking only at improvement?

(Ann Beach): Yes. So I mean we do see decline in patients. It does - there's not that many but we do see decline in function for people who have, you know, are admitted in the middle of the scale or, you know, near the top of the scale. In terms of, you know, the - so those patients are included in the other measure where basically we're modeling where somebody could be going - having improvement or decline.

In terms of these particular measures we exclude people who basically have these incomplete phase. That's where you potentially see the most decline in function.

But as I said the main - the reason for being admitted to an inpatient rehab facility is functional improvement and so the vast majority of patients do show improvement and we really don't see people really at the top of the scale. It's very few who would be independent in all those areas.

(Brenda Lee): Can I interject here? We're excluding the person from the self-care measure who's independent at the time of admission.

(Ann Beach): With the self-care item.

(Brenda Lee): Right. And I know we're talking about both of these measures kind of at the same time. If it's a mobility measure I would expect you got a lot of patients who are admitted to a rehab for mobility. It's the classic for hip replacements. And they have all their marbles and are appropriately, right.

So maybe what's bothering (Ellen Sholts) and me a little bit is the sort of separating people into pieces again. That we do think that there's some validity in taking a look at whether or not somebody functions have declined.

((Crosstalk))

(Ann Beach): (Unintelligible).

(Brenda Lee): Yes.

(Ann Beach): CMS in terms of their coverage requirement in order to be admitted to an IRF you must need care from at least two disciplines. And often it is OT/PT, sometimes speech is also included.

So I mean people do have self-care deficits and self-mobility deficits in order to be admitted an IRF.

(Ellen Sholts): (Unintelligible) and my…

Lee Partridge: Yes.

(Ellen Sholts): ...(unintelligible) and you have somebody that's had a stroke and their mobility is intact but their ability to do some of the other self-care activities are not intact because of the stroke. And maybe they need to do some language and they need some OT for addressing cognitive training. So that's just an example. Or, you know, you can flip it the other way where you could have someone who, you know, anyway. Sorry, I don't want to take up too much time.

But there are ways that you could imagine that having these as separate measures would make sense immensely.

Lee Partridge: Okay, any other comments about validity of the measure?

Okay, let's vote on accepting or not the Scientific Methods Panel. Again they voted moderate on average.

Woman: Okay voting is now open for Measure 2633 on accepting these SMP ratings for validity. The options are yes or no.

Lee Partridge: (Unintelligible). I heard two more, one more. Okay (unintelligible).

Woman: Voting is now closed for Measure 2633 on accepting these SMP ratings for validity. The majority vote is yes with 17 votes and 2 votes for no.

Lee Partridge: Okay, now we're up to feasibility. (Brenda) do you have any comments or other discussants have any comments about feasibility?

(Brenda Lee): No major comments from me other than, you know, they're using standardized data only so and they have been doing this - they have been using standardized data elements. And they've been using these measures over a period of time.

So in my view it's feasible for them to use the measures. So I don't have any other comments beyond that.

Lee Partridge: Okay, how many discussants? (Ellen Sholts) your card is up. Did you have a thing or is that from before? Okay. There were no other concerns from the pre-evaluation. Okay, should we vote on feasibility?

Woman: Voting is now open for Measure 2633 on feasibility. The options are A, high; B, moderate; C, low; and D, insufficient.

Lee Partridge: Waiting for two more or not.

Woman: At least (unintelligible).

Lee Partridge: One, two (unintelligible).

Woman: (Unintelligible).

Lee Partridge: Okay.

Woman: Yes.

Lee Partridge: So we're good.

Woman: Voting is now closed for Measure 2633 for feasibility. The criterion does pass with 7 votes for high, 11 votes for moderate, and 1 vote for low.

Lee Partridge: Okay usability and use.

(Brenda Lee): So in the documentation the developers have indicated, you know where the data is - the measures have been reported publicly.

And my understanding is that this year you will have something else but you'll be reporting it out on the CMS web site. Is that correct?

(Ann Beach): The data from Calendar Year 2019 will be publicly reported in 2020. So public reporting has been finalized through rulemaking and will happen next year on IRF detail.

(Brenda Lee): Okay. I don't have anything else.

Lee Partridge: (Don).

(Don Casey): I don't know if this is one question or two. But given the two year period of no change what are the possible - what are your possible explanations for that? And, you know, is one of them the possibility that it's not clear to the end users how to apply the data to make change?

(Ann Beach): Let's see. I think there's - I guess it's a couple of questions. So one thing as I mentioned before there has been a lot of changes going on in patient rehab facilities and also acute care in the last couple years because of the IMPACT Act so getting used to new data elements, thinking about the future, is definitely a part of I think, you know, people have been focused on making sure they're requesting accurate data and, you know, coming to all the training programs, understanding what's going on with the measures and kind of the plan for the measures.

You know we do see differences in facilities. We definitely do see that. In terms of why there wasn't change, you know, I think there's a few things that might be going on.

But one of the things I think we worry about is that, you know, when the data are used for payment starting in October 2019 we hope that people will not change coding practices. Because basically the admission scores just like the FIM admission scores are used for payment. And so, you know, there's an incentive for people to code people lower and we wouldn't want to see changes relating to coded practices.

And I, you know, I personally get emails from software vendors saying we know how to help get your FIM scores lower so you can get paid more.

So we would not want that to happen. And so we will be monitoring and in fact CMS asked us to monitor the data very carefully to make sure that the data do truly reflect what's going on and don't show changes.

Alan Levitt: And I also think the question is one year enough time to really see such changes. I mean I'd certainly want to continue to monitor. And the hope is that in the perfect world where hopefully providers and stakeholders don't have other motivations for coding that would actually change the measure itself because of external motivations that it won't influence the measure score.

But our hope is that the provider scores will improve and we'll see improvement in both the change and in the percent needing.

(Don Casey): So but I guess what I'm hearing is we all agree we're speculating and hoping and you don't have hard facts about that. And one possibility is it's hard to use that score to make change, one possibility, one speculative possibility. That that is a possibility, correct?

Alan Levitt: Well I guess if I was a witness under oath if it's possible.

(Don Casey): Well no, I'm just asking because I'm concerned…

Alan Levitt: No but, no I mean, right.

((Crosstalk))

(Don Casey): …about this (unintelligible).

Alan Levitt: I mean I think there are a number of factors that can influence scores. I mean we certainly have seen in the - that performance scores or other measures like the measures that are using FIM that there are ways that providers (unintelligible) use that are external to any sort of change and actually have a patient to actually doing and, you know, those are shown in terms of both lowering of the admission score. I guess it was called underestimating of the admission score and a bow effect and use of independent days at the end.

So there are different strategies that are used. And certainly our fear is that certainly our items we want to try to prevent that and to have a data validation strategy to be able to identify that so that any sort of changes that we truly see would be changes that are true to actually improvement in function.

(Don Casey): But it's not unreasonable for you to accept my premise that there is a possibility that these (unintelligible).

((Crosstalk))

Alan Levitt: Always, right.

(Don Casey): Okay.

Alan Levitt: Right. You know, just understand, I mean we will be continuing to look at all of our measures. You think you're a bad critic. We're the worst critics of our measures and…

(Don Casey): I'm not being a critic.

Alan Levitt: No. No, but I'd just say, we are. And we will continue. And if we see things such as that we will continue to look at the measures and say what can we do with these measures to make them better?

And we will always be coming back here with these measures. You'll see continued changes all the time because that's what we should be doing.

(Brenda Lee): I just wanted to note that this conversation underscored something we've been talking about a little bit earlier today. Right now this measure is used for quality. In October it moves over to become a P-for-P and the stakes (shapes).

(Ann Beach): PPF, PPF not pay for performance, just PPF.

(Brenda Lee): That's correct.

(Ann Beach): There's no pay for performance right now in our…

(Brenda Lee): Right (unintelligible) but…

((Crosstalk))

Alan Levitt: Again it's the…

(Ann Beach): Yes.

(Brenda Lee): Yes, it's moving to a different thing, different sentiment, different...

Alan Levitt: The items behind the measure will be used. The measure itself is not used for…

(Brenda Lee): Right.

Alan Levitt: For payment. It's the items that are there or some of the items that are there.

(Ann Beach): Yes. And I'd also like to state that in terms of the items, the original work that went into developing the items the goal was to actually measure a wider range of function because they were intended for cross setting use and so they were to be included data elements for long-term care hospitals where you have the very chronically critically ill patients.

And so there's a fair number of bed mobility activities in the mobility measure and then in the area of home care the items are also implemented now.

And so again the goal was to have this wider range of patient functioning to measure. And so we have done some analyses comparing some data. And I'm not sure we shared all of this. But we could certainly share that.

Lee Partridge: (Ellen Sholts).

(Ellen Sholts): So and it occurred to me listening to this debate that as awkward as the situation is to have competing measures today it's also an opportunity where you could test this question empirically, right. I mean we have in front of us here what the change in score and the FIM-based measures for Fiscal Years 2017 and 2018. And we have here the change in scores for Fiscal Years 2017 and 2018 for your measure.

So it is possible to look and you could monitor this over time. You have a very similar measure for the same or very similar populations that you could look at the level of variation you're seeing between the two measures.

And if you're seeing something different between the two of them it's suggests that, you know, like one more is more sensitive than another or less reliable or particularly if you see a change between now and when payment gets tied with these that there is some sort of (scheming) or adjustment in the coding that's going on so some food for thought.

Lee Partridge: Excellent. I believe (Tom), (Ann).

(Brenda Lee): Quickly, how many IRFs are there?

(Don Casey): Eleven hundred.

(Brenda Lee): How many? Eleven hundred and fifty.

(Don Casey): Eleven hundred and fifty. About 350 free standing and the rest are units in acute care hospitals.

(Brenda Lee): Acute care hospitals or nursing homes.

(Don Casey): No.

(Brenda Lee): Might say be a unit of a nursing home.

(Don Casey): No.

(Brenda Lee): No.

(Don Casey): It's a hospital level of care.

(Brenda Lee): It just strikes me that's a really small number. I don't know what I thought it would be but 350 standalone for all of this measurement just seems…

(Don Casey): Well there's about 1150 total approximately.

(Brenda Lee): Thank you.

Lee Partridge: Great. These - couple really good points made just in the last few minutes. Good. Anything else about usability and use and keep some of those points in mind too because when we talk about the competing measures that'll be relevant.

(Terry): Just a more general comment. Just to - come back to the days prior to the IMPACT Act when the four federally mandated assessment in some of the acute care sites could not be cross walked. They're all different. They all had different measurement concepts, different source and ceilings on the measures. And you couldn't measure someone's performance on one side of care to another.

And I would add this to the usability piece and that becomes just a really critical issue in creating episodes of care in measures that will cross episodes. And this is really sort of building (unintelligible).

(Brenda Lee): The fact that we can't do that I don't - maybe (unintelligible) Mr. (O'Malley). No, what we're - what's his name?

Woman: (Terry).

(Brenda Lee): (Terry) sorry. Are you saying that it's a problem that we can't measure them across sites because the difficulty in episodes of care or were you saying we

need to be able to measure them differently in different sites? I just didn't hear you clearly.

(Terry): I'm sorry. I probably wasn't near the microphone. It's we couldn't measure them across sites of care because the measurement instruments are all different. They measure different things.

So when someone left a hospital and went to a SNF or went to an LPAC and then went to a SNF and they went to home care, they all measured different things. That's changed now with the IMPACT Act. And Section GG which is the standardized shared assessment instrument.

And that's what's going to drive function, cognition, pressure ulcers.

(Brenda Lee): (Unintelligible).

(Terry): Yes. And so the real goal was from a patient's perspective, what was my experience across my episode of care. And if the results occurred in several different sites, that's not possible currently but will be.

(Ann Beach): Right, so just to build on what (Terry) said so this Section GG data elements that we're talking about for the IRF has been implemented in IRFs, SNFs, LPACs and home health. The home health was the last group to implement in January 2019.

Lee Partridge: Okay. Good, any other usability, use points before we vote on them? Okay, well let's go to vote.

Woman: Voting is now open for Measure 2633 on use. Your options are A, pass; or B, no pass.

Okay, voting is now closed for Measure 2633 on use. The criterion does pass with 15 votes for pass and 4 votes for no pass.

Lee Partridge: Good and let's proceed right to usability because I think we're assessing them both at the same time unless there are some other points. Okay.

Woman: Okay voting is now open for usability for Measure 2633. Your options are A, high; B, moderate; C, low; and D, insufficient.

Okay voting is now closed for usability in Measure 2633. The criterion does pass with 2 votes for high, 12 votes for moderate, and 6 votes for low. And that's 70%, so 10% for high, 50% for moderate and 30% for low.

Lee Partridge: Okay and overall suitability then.

Woman: Voting is now open for the overall suitability for measure 2633. Your options are A, yes; and B, no.

Voting is now closed for overall suitability and endorsement for Measure 2633. There's a unanimous vote for yes with 20 votes.

Man: (Unintelligible).

Lee Partridge: Okay great. Okay. So we'll plow through and do the mobility measure. Hopefully we won't need as much discussion time and then we'll take a break after that. We promise. Hand it over to (unintelligible).

(Susan Sparcet): So moving onto 2634 change in mobility score. We have four lead discussants - four discussants. (Ann) has the lead. And (Linda Murillo), (Sherry) and (Brian) are also - have also been reviewing this.

So (Ann) do you want to lead us off and are there issues?

(Ann Beach): (Unintelligible).

Lee Partridge: Well.

(Susan Sparcet): They have dealt.

Lee Partridge: They combined this.

(Susan Sparcet): Yes.

(Ann Beach): (Unintelligible).

(Susan Sparcet): Sorry. And they're related to - just I need a little clarity. This measure is only for people enrolled in Medicare or Medicare Advantage…

(Ann Beach): Correct.

(Susan Sparcet): …which to me narrows its usefulness in a setting with a lot of different payers involved. If this is approved, endorsed, and they want to expand it to other populations does it need to come back through the process to get endorsed it for different populations?

Woman: Yes. So you should be looking at the specifications (unintelligible) the intent of the measure. So if they wanted to expand the intent the specs need to be revised. And they agree as well, (Ann)'s shaking her head.

(Susan Sparcet): Well I would just like to say that I hope that they start on that process very quickly because I think this is another bifurcation of a population within a setting that puts a lot of - could put burden on the organization and also mask overall organizational results.

So I just want to say that. I did have a question too. In the numerator and denominator discussion you used the term "There's no simple form of numerator and denominator."

And I don't know what that means. Can you help me understand?

(Ann Beach): Sure. So basically the measure is change in mobility. In this case we're talking about mobility measure. And we actually calculate an observed change in mobility for everybody, all the patients in each facility.

And then we look at - we create based on the risk adjustment an expected score and then we look at the difference between the observed minus the expected score. Look at that difference and then add on the average national mean change in mobility.

And so there's not really a numerator/denominator like there is in the 2635, 2636. It's basically a calculation of the change and then it gets risk adjusted. And there's not a numerator/denominator.

Lee Partridge: (Unintelligible).

(Susan Sparcet): Okay. I'm still not sure I understand but I'm sure my colleagues will help me with that. I don't know if you want me to go into all the various pieces, evidence, reliability, or I should wait until we come to them, (Chris) or Lee Partridge how you wanted to do that.

(Chris): (Unintelligible) colleagues participate in a general comment.

(Susan Sparcet): Right.

(Chris): (Unintelligible).

(Susan Sparcet): Fine with me.

(Chris): So I found this measure to be very simple and straightforward. One of the things that I have found with some of the other metrics of rehab use is that they can be difficult to score.

And echoing the comment that was made by the one public member who had said that it helped her to understand what she needed in place or what caring for that person at home would be like because she understood what the measures were.

And I spoke again with one of my colleagues and she confirmed. These are much more precise, easy to understand metrics. And that that makes them very useful for a variety of things and that being said, you know, there are some concerns. And one of them being the population limitation but other than that was all I had for now.

Lee Partridge: (Unintelligible) thank you. I agree with that. I found that to be important.

And I guess just one note that kind of leaks into the evidence piece but the fact that there were studies that demonstrated that a year after discharge there was higher mode of function I found to be very important. This is exactly the kind of thing that policymakers want to be knowing about in order to make decisions about how to spend healthcare dollars.

So but that's all I had as open comments.

(Ellen Sholts): Yes. And I don't have anything to add in terms of comments but just a question to clarify, so for this measure the decisions do not include patients who come in for mobility. Because you said before on the previous measure that they came in sort of at this highest score for self-care that they would be excluded.

(Ann Beach): You're correct, that the exclusion criteria say that if somebody is independent in all mobility activities they'd be excluded from the measure but…

(Ellen Sholts): But somebody…

(Ann Beach): …somebody's who independent going up and down 12 steps and all the bed mobility activities probably, you know…

(Ellen Sholts): Right.

(Ann Beach): …wouldn't be qualified for…

(Ellen Sholts): So they are excluded. I just wanted to clarify that.

(Ann Beach): They are excluded. Yes.

(Ellen Sholts): Okay.

(Ann Beach): They probably wouldn't need IRF level care.

(Ellen Sholts): Yes.

(Don Casey): Yes that's right.

(Chris): Yes, right.

(Don Casey): They probably - they didn't in the first place.

(Ann Beach): Exactly, exactly.

Alan Levitt: Can I just take a second?

(Ann Beach): Then why exclude them?

Alan Levitt: Oh I'm sorry. I just wanted to address the population question because it keeps coming up. I have a saying when it comes to measure development that you can't always make the cookies with the ingredients you have in the pantry.

And we are limited by, you know, what we're able to collect. And so, you know, the measures that you see we will be collecting data on everybody we possibly can. And that's the reason why again in our fiscal year we will - that proposal can - we, you know, propose to expand that.

So that particularly when we're looking at these measures that are Impact Act measure that are involving these standardized assessment data. That are going

to, you know, be looking at the sorts of things that could be looked at the core setting. That we want to have apples to apples types of comparison in terms of the patient population.

Man: I had this question before but just kind of give you an option here to just kind of address it. And that is the difference between claim 2634 and 2636 which I'm about to talk about in a moment. Why both?

And what are they designed to achieve? Different from one another. Is one more susceptible to gaming than the other?

Woman: Excellent question. So the reason for the two sets of measures is based on some research that I did about 10 years. Where we actually - we made fake report cards and asked, you know, generally public people, normal public people how to interpret quality measures.

Including change in self-care, change in mobility. And the consumers had a great difficulty understanding what that meant. It wasn't a normal thing.

Like and they were like well what does this mean? Like, how much should somebody gain? And even thought we had, like, a national made up.

It was all fictitious here we did for the project. But they really didn't know how to understand. But they did really understand measures that had numerators and denominators.

Per so the 2635 and 2636 are measures where you actually report percent of people who met or exceeded the national average based on (Rick's) adjustment. So percent of people - you have a benchmark. And consumers understand those types of metrics.

They can interpret them quite easily. Whereas with the change in measures that's something that actually the (ERP) are very used to. It takes advantage the granular data that you have.

Every unit of improvement is actually reported there. And so we didn't think that the providers wanted to just have a general percent of people who never exceeded the benchmark. Because we lose a third of the data in that.

So we basically, you know, if we could we would have combined them into one measure. But you can imagine the forms would have been a nightmare to understand if we did that. And so we ended up actually having two separate measures. Does that help?

Man: Yes it does. And in terms of the risk adjustment we can get into I suppose during my discussion - the discussion for 2636. But those risk adjustments - in fact can you tell me a little bit about the granularity of those.

And what do they really capture. Because those - that measure really only makes sense or is really useful if you've got a target that is useful - that's accurate, right.

Woman: Right. So the risk adjustment co-variant are the same for the two self-care measures and the two mobility measures. In general, you know, our process was to review the literature to see what the literature said about what our patient factors associated with functional outcome.

So that was our starting point. We had several expert panels. The list of the experts are in your materials.

And then we actually did data analysis after that. So it was basically this process where we looked at evidence, we talked to experts, and then we did the data analysis and said what factors are actually associated with functional improvement. In general many of the - like the missions self-care was significant obviously for the self-care measures.

The mission mobility was significant for the mobility measures. We actually found that wasn't a linear relationship so we include both the continuous form of admission self-care as well as the quirk version. We looked at (comovidity).

We used the hierarchical conditions categories which CMS uses for risk adjustment. And so we basically went through everything that we could in terms as you said to try and find and adjust for everything that made sense clinical that could be adjusted for with the data that we have.

Man: Great. Thank you.

Woman: And (Anne) I think it's time to read awkward evidence.

(Anne): Oh great here I go. I thought it was interested the question that we were supposed to ask ourselves about is there one thing that will improve your performance on this measure. And there was a discussion somewhere about for this measure really to work there has to be individualized plan for every person.

Which I agree with. But doesn't leave you with kind of a commonly implemented intervention. (Ryan) talked about the follow-up study, one year after. But I assume that was with a different tool, a different measure.

And I'm wondering if you have , is it a second measure to do a one year follow-up or six month follow-up. Or is that baked into this measure for use.

Woman: So I believe the follow-up would have been a - the literature that we sited.

(Anne): Yes you did. But…

Woman: Yes.

(Anne): …I'm asking about your implementation of this measure and if you see follow-up six months, one year whatever later…

Woman: Yes.

(Anne): …part of the life of this…

Woman: Yes.

(Anne): …achievement.

Woman: So this measure is actually change in function between admissions and discharge in (ERP).

(Anne): Right.

Woman: There's no follow-up component. I agree, you know, it would be nice to have that at some point. But the data are actually collected by clinicians.

Its clinician reported data and because follow-up tends to be, you know, not everybody replies to surveys. You know, it's really just admission discharge after everybody.

Man: And again this measure's based on the mandate that the Congress passed on the Impact Act between admission and discharge. And so that's why you see these two time points for this measure.

Man: Follow-up could I just ask a quick question about the patients that have home care and (unintelligible) you would have data on them, right? Maybe analyze that because the follow-up thing to me is key, right. What happens next?

Woman: So we do have that data. We haven't had a chance because we've been working on these forms to actually track people over time. But certainly that is part of what (EMF) is interested in doing. Exactly.

Man: As Dr. (O'Malley) said I mean the whole goal of these items is to be able to be used longitudely in the care of patients in that we would be able to take those items and use them across healthcare settings for measures such as this. And be able to use them as well to determine resource use as well to help in terms of payment. Multiple purposes.

Woman: I do want to talk about disparities but I think I'll wait. Is that - that's a gap issue. Okay. My colleagues.

Woman: Any other comments on…

Woman: Yes I just wanted to add that I also like to see if there's a way to transition to usage of measures like this in outpatient settings as well. But that - especially in rehab that's a very significant portion of the care.

Woman: Okay. There are no further comments or discussions we'll move on to taking a vote. And 1A (Nadia).

(Nadia): Voting is now open for measure 2634 for evidence. Your options are A for pass and 2 for no-pass - B for no-pass.

Woman: I probably lost my thought.

(Susan Sparcet): The link is in the calendar…

Woman: I got it.

(Susan Sparcet): Okay.

Woman: Thanks.

(Nadia): Okay. Voting is now closed for 2634 for evidence. And it is a unanimous vote for 20 votes for pass.

Woman: Okay moving on to GAP 1B.

Woman: Well clearly there's a wide gaps on a variety of social economics and demographic rates. Like the other ones. Did you look at urban versus rural?

I live in a relatively rural area. And we see these disparities magnified in those areas. I'm less concerned - I mean so there's clearly a gap to be measured.

And I think it's a - it becomes a real issue in payment and accountability that you said we're going to be doing in September. And there's such a large gap. But I think - so you certainly shown a sufficient gap. I just worry about moving it too quickly to build it in to revenue for - and payments and deductions for providers.

Woman: Comments on gap? No. All right vote on 1B please.

Woman: Table 14 if you want to look at the data.

(Nadia): Voting is now open for measure 2634 for performance gaps. You're options are A-high, B-moderate, C-Low, and D-insufficient. Voting is now closed for measure 2634 for performance gap. The criteria does pass with 6 votes for high, 12 votes for moderate, and 2 votes for low.

Woman: Moving onto our good friends, scientific susceptibility. We have recommendations within the (unintelligible).

Man: Yes we do.

Woman: Okay.

(Susan Sparcet): High for reliability and moderate for validity in the (unintelligible) account.

Man: Unintelligible.

Woman: Okay. Are there any discussion? (Anne)?

(Anne): Not from me.

Woman: All right. (John)?

(John): No.

Woman: Okay. Then we would move on to vote on accepting the recommendation of the panel on reliability.

(Nadia): Okay. Voting is now open for accepting the S&P's rating for reliability for measure 2634.

Woman: (Unintelligible).

(Nadia): Okay. Voting is now closed for measure 2634. All the Committee members have voted yes for accepting the S&P's rating for liability.

Woman: Moving on the second half accepting the panel's recommendation on validity which is matter.

(Nadia): Voting is now closed for measure 2634 for accepting S&P's rating for validity. And the Committee does vote to uphold the S&P's rating for 15 votes for yes. And 4 for no.

Woman: Okay. Feasibility (Anne).

(Anne): No comment.

Woman: We're getting tired out.

Woman: (Unintelligible) late.

Woman: Anybody else. (Tommy). All right moving on to voting on feasibility.

Man: There's something (unintelligible).

Woman: I agree.

(Nadia): Voting is now open for measure 2634 for feasibility. Your options are A-high, B-moderate, C-low, and D-insufficient. Voting is now closed for measure 2634 for feasibility. And the criteria does pass with 9 votes for high and 10 votes for moderate.

Woman: Moving on to - I always forget the…

Man: Use.

Woman: …user usability. I forget what comes first.

(Anne): I do have a couple comments on that. I'm sure that you're - the measures develop the way it is with the population that it is because of the mandate that you have. And my guess is along with that mandate comes a mandate to collect data as well. So this is going to be implemented I'm assuming given where we are.

And I'm just concerned when it's implemented for payments and accountability that you just keep your (unintelligible). The other unattended consequences of something moving to baker performance too early, I think we at least in New York have seen it time and again cause exactly the opposite behavior that you want. Facilities closing, no place for people to go. So - but I respect the idea of a mandate, I urge you to keep your eyes on this very closely to keep it from becoming a punitive system.

Man: Okay. Thank you for comments. Can I respond?

Woman: You can.

Man: Okay. Thank you the measure was adopted and is, you know, currently being collected by IRF. So they're already - they are collecting this data and as we mentioned the data will be publically recorded for the changed measures in 2020.

Right now the - all of (IRF's) in this case are required to submit the necessary data. Or they're subject to a 2% reduction in their rates. So some people call it pay for recording.

If there's a pay - adopt pay for non-recording. But anyhow it is not tied to any value-based purchasing program. Any decision that would ever be made regarding such a program in the IRF setting or in any setting would be coming from Congress not from CMS.

We would end up implementing that program if Congress asked for such legislation and they asked that this measure be included in that program. Once again we would implement that measure. But none of that has happened yet.

Woman: Well maybe we could have a little offline discussion about it.

Man: Okay. No…

Woman: (Unintelligible)

Man: …I understand.

Woman: (Unintelligible).

Man: Right. And again as we mentioned before take the measure out it's really the items as well. Because, you know, as (Anne) mentioned and has been mentioned around here before.

These are items that are used and planned to be used for multiple purposes. Including some of them being used for resource use for payment. So even if there's no pay - a value-based purchasing piece to it. Certainly we are concerned that there may be other motivations for the coding of such items that could involve both improvement in their measures scores and also changes in their payments.

Woman: I'm beginning to see, all though I've been guilty of complicating the issues here. But I think we need to unpack a tiny but - bit. What - we've seen references throughout the materials to this - these measures for being used for a perspective payment systems in Medicare. I think what we need to understand is what that means.

Man: None of the measures discussed today are used in perspective payment system. The items that were discussed previously were used as part of payment rates that are - were being assigned to IRF. And…

Woman: In other words they were used to determine the severity, the intensity of the care. But it's almost like what we do in the hospitals. We're…

Woman: (Unintelligible) admissions.

Man: Right the admission items would be used to help to determine the resources necessarily for the patient within this particular setting. Correct.

Woman: And what we're talking about now in the context of Medicare is the same function. Different function?

Man: Well what we're talking about right here is a measure that's being adopted and used in a quality reporting program.

Woman: Right.

Man: So it's just used for public recording.

Woman: Right.

Man: It's not used for any payment purposes.

Woman: Does that make sense to you? You live in this world more…

Man: It does make sense to me. But I don't think anyone believes that that's not where we're headed. I mean that's clearly where we're going in the future.

However that plays itself out. So I appreciate that that's not today's circumstance. But that's - is it's past it's time. I fully - I think every - most policy makers expect that that's where we're headed.

Man: Yes. And again if that occurred, and if the legislation was, let's say similar to the legislation that been used for all of these programs, any measures that would be considered for such a program would likely come to a Committee

like this for, you know, consideration. I mean I can't be absolutely certain of that. But again that's been done before.

Woman: You're right.

Woman: (Unintelligible).

Man: No -

Woman: (unintelligible)

Man: Can I go ahead? Thank you. I just wanted to point out that the evidence that comes in from a measure like this could also support providing a certain level of care when someone might be proposing to reduce that care. So it could work the other way around in a positive way for beneficiaries.

Woman: So denials and so on.

Man: Yes. My question simply was do you know what IRF's are doing - I hate acronyms. Inpatient Rehab Facilities are doing with their non-Medicare patients now that they're forced into this arrangement?

Woman: Are you asking…

Man: Do you know how, if, when, they're collecting data using you're method or others. Do we know what happens to the rest of their populations? Visa via measurements of these issues.

I'm curious. And if we don't know, you know, maybe we should find out. I mean this maybe another kind of consequence of them saying the heck with it. I doubt it. But I don't know. (Paul) (Unintelligible).

Man: I - every IRF that in know usually defers once the Medicare standard is established. They just do it for all patients because it's easier. And most payers require something. And they typically will just look to Medicare as the standard.

Man: Thank you.

Woman: Right. And just a reminder as we said before CMS actually has a proposal out before to go to all payer for the IRF pie. Which, you know, potentially opens the door in the future.

Woman: Any further discussions on use. All right. Voting (Nadia).

(Nadia): Voting is now open for measure 2634 for use. You're options are A for pass and B for no-pass. Voting is now closed for measure 2634 for use. It's unanimous 20 votes for pass.

Woman: And finally usability. Any comments or discussion about that? No.

Woman: Yes I just wanted - I'm not sure how much the committee knows. But typically these metrics are - the way their used in facility is that there are a number of people who provide input on what the measure with the rating is. So you have nursing aids, nurses, along with clinicians who are giving information about somebody's mobility.

Say for example, because they're the ones helping you to the bathroom and so forth. So a simpler metric is easier to gather that information on and how to be reliable I would think. But it's certainly simpler to gather. So in terms of usability within the facility I think it would do well.

Woman: Thank you. Any other comments? All right. Vote (Nadia).

(Nadia): Voting is now open for measure 2634 for usability. You're options are A-high, B-moderate, C-low, and D-insufficient. Voting is closed for measure 2634. And the criteria does pass with 6 votes for high, 11 votes for moderate, and 3 votes for low.

Woman: Final vote on recommendations.

(Nadia): Voting is now open for the overall usability for endorsements in measure 3634. You're options are A yes and B no. Voting is now closed for measure 30 - 2624 for overall usability for endorsement. And the Committee recommends measure 2634 for maintenance of endorsements with 20 votes for yes.

Woman: Okay. Thank you all very much. We need a break.

\

We need to walk around. Maybe turn cartwheels. I don't - come back for our final and toughest session.

In how long 15 absolutely. Yes 3:05.

(Susan Sparcet): Yes back at 3:20.

((Crosstalk))

Woman: The group.

Man: No I mean the meeting is…

Woman: it went well.

Man: That was an unusual (unintelligible).

Woman: I guess and it helps to have gone after the one before.

Man: Right. Now they put mine off those calls…

Woman: Yes.

Man: …but that's a newly (unintelligible) age gap is very…

Woman: And I have the health funds gap because of church users.

Man: Oh yes.

Woman: I'm, you know, not the lead on that.

Man: Right. But very close.

((Crosstalk))

Woman: (Brian) you got to get up.

(Brian): I should huh.

Woman: Yes at least stand for one minute. Stretch your legs.

(Brian): I will okay. I'm going to do it.

((Crosstalk))

Man: Well it's crazy. (Unintelligible) we're actually puzzled (unintelligible). It's terrible.

Woman: (Unintelligible).

Man: It is - so (unintelligible) function is not (unintelligible). It is really - we're statewide in Washington.

But they wouldn't (unintelligible). Yes maybe so far. Well we're grateful. Wow.

Woman: (Unintelligible).

Man: Yes.

Woman: (Unintelligible).

Man: Oh yes. One dos a lobbyist for that.

Woman: Well they (unintelligible) multiple (unintelligible). Yes.

Man: (Unintelligible) the outbreak was.

Woman: That's right.

Man: (Unintelligible) outbreak.

((Crosstalk))

Man: It varies a little but (unintelligible) seven (unintelligible). If you were born within the 70's (unintelligible).

Woman: Before (unintelligible)

Man: There was so much you all had to do. You were percolated and that's (unintelligible).

Woman: Yes.

Man: And then when my (unintelligible).

((Crosstalk))

Man: How much do you (unintelligible).

Woman: It's nice to be (unintelligible).

Woman: Oh awesome. Isn't that pretty?

Woman: I know I'm like trying to (unintelligible). Supposed to be right back here. And so (unintelligible).

Woman: Oh cool.

Woman: (Unintelligible).

Woman: (Unintelligible) for 15 sweets.

Woman: Wow.

Woman: For embassies and all that. And everything else. But it's so funny…

Woman: I know. Yes (unintelligible).

Woman: You can get the (unintelligible) done.

Woman: Yes. (Unintelligible). So cool.

Woman: I'm trying to think what it is.

Woman: Oh I thought that was…

Woman: …(unintelligible) that's as far as we got.

Woman: We probably can (unintelligible).

Woman: Yes. I need another (unintelligible) today.

Woman: That's yes. I'll see you soon. Say hi to your boys.

Man: I'll do that.

Woman: Yes it's just…

Man: Okay. These are great discussions. But we do have to finish.

Man: Okay.

Man: I know all of us would like to be finished right now.

Man: Thank you.

((Crosstalk))

Woman: What are we going to do? Invite everybody to (unintelligible). Are we there? Okay good.

Man: Okay.

Man: And share.

(Susan Sparcet): Okay. Thank you everybody. We are now going to have the competing measures discussion. I'm going to turn it over to (Alisa) to make a few welcoming remarks before we talk about profit.

(Alisa): So thank you everyone. And thank you not just for today. It's been a great discussion but thanking the Committee and developers.

UDSMR, and (CMS) for some pet discussions over the last three years. I've been at NTF for almost 10 years. It'll be 10 years in January.

And what we did 3 years ago is almost unprecedented at NTF. What we tried to do in evaluating the measures is to make that we're putting out measures

that have strong scientific rigor. You know, hoping that the measures are not creating burden as they go out.

And hoping that they will be used for improvement. We want to make sure that as we're evaluating measures we are not creating burden as we're doing so. And so one of the criteria we don't talk about often because we often don't see this in our work.

We may see a lot of the laden measures. But not measures. We had the interesting experience about 3 years ago to have the 3 sets of competing measures.

And at the time our Board of Directors was responsible for ratifying endorsements. And this Committee worked really hard and struggled because these measures were so similar that they could not make discussions in terms of which ones are best in class. So the Board also was, you know, responsible for making the ultimate endorsement decision.

They couldn't often decide which measures were best in class. And they had hoped that through the maintenance, the 3 years we've been in, that we would learn information about the measures. About the implementation, the use, the unattended consequences, everything that we've been talking about today.

I think that today we learned that that was challenging as well. And, you know, they are a lot of positive attributes to both of these measures. But I just wanted to set the stage a little bit.

Talk about what we require (unintelligible). This is not part of our process of what we intend for measures as we endorse them or recommend them for

endorsement. We want to be able to determine which are the best ones to be implemented.

But I also want you to know that we recognize that these are hard decisions. We tried to bring you as much as information as we can get from the developers we've worked with over the last 3 years. And some other end users as they've had experiences with these measures.

And so with that I will turn it over to (Van) who will walk us through the algorithm. So we can add this to our discussion.

(Susan Sparcet): Okay. Thanks (Lisa). Could you pull up the next slide with the measures? So we are here to talk about two steps of competing measures.

Of course 2286 versus 2633. And 2321 versus 2634. Next slide. Just briefly this is NCF related versus competing (decisiontry).

Same. So we're looking at two sets of measures with the same pertinent population and the same measure focus. And then the process, I think we've actually already discussed this.

So you can kind of skip ahead the next couple slides. And we're actually focused right now on figure 1 which I think is probably a bit hard to read. Oh sorry - let's - yes let's talk about the history briefly.

I think (Alisa) touched on it but just we're looking for a decision and we did incorporate a lot of the previous discussion and memo in the memo that we sent out so that you have that to hand. And then figure 1 which is up on this main screen here and also in the memo in a more readable format I expect. Helps you walk through how NCF sees the process working.

We have already gone through the first step of does the measure (unintelligible) the evaluation criteria. And so now we want you to work through the bottom. Next if you could scroll down to Table 9. Yes.

So with the functional change, change in self-care score. That's 2286 from UDSMR versus 2633 inpatient rehab facility functional outcome measures change in self-care for medical rehab patients. So the first UDSMR measure and the first CMS measure that we talked about are competing with each other. And then the second, the other pairs 2321 functional change, change in mobility score from UDSMR compared with 2634 inpatient rehabilitation facility functional outcome measure change in mobility for medical rehab patients from CMS.

Man: (Unintelligible).

Man: No we're not.

Woman: We're not. We're comparing - so we're still not comparing 2633 and 2634. No 26 - 22- let me get the - 2636 and 2635.

(Susan Sparcet): Those are considered relating - related…

Woman: Okay.

(Susan Sparcet): …not competing by NQS criteria.

Woman: Okay.

Susanne: So we're not talking about those. We will talk about those separately after - originally we had planned to talk about all the CMS measures together. But because they are not we want to have this competing measure conversation in person. We're going to talk about the competing measures first.

Woman: Okay. To the fact when we voted on 2, 15 it was 2636 and 2635 that were the competing CMS measures. That we voted on, right in 2015. They're the ones that are up for renewal.

(Susan Sparcet): 26…

Woman: The ones that we voted on today were new. We hadn't seen them before.

Man: Yes we have.

Woman: Had we? The ones…

Woman: Yes.

Woman: …we did today. Two, six, three, three and 2634 are old. We were renewing them? I thought they were new. Okay. All right.

Man: So this is a prior discussion. Discussing…

Woman: Okay.

Man: …them. Unfortunately we were interrupted. At the time did not arrive at a conclusion.

Woman: Yes.

Man: And some of the…

Woman: Yes.

Man: …rationale is spelled out with the slide. It really had to do with the fact that we just couldn't decide on the best in class. So we elected actually against our own criteria to move forward with both sets of measures…

Woman: Yes I remember.

Man: …to endorse.

Woman: I remember that the session is to…

Man: Yes.

Woman: …again 2636 and 2635 were so similar to these measures.

Man: Right.

Woman: There's a little bit of methodological difference in how you derive the numerator. But, you know, they're…

Man: Yes.

Woman: …fundamental - from a consumer's perspective there's not a huge difference.

Man: Yes. Can't blame you for…

Woman: Right.

Man: …making the (compilation). They are remarkably different.

Woman: Yes.

Man: Sorry to interrupt. (Susan) one more time. There's four numbers involved…

(Susan Sparcet): Yes.

Man: …we're comparing two two's.

(Susan Sparcet): Yes.

Man: Can you just quickly go over those numbers one more time so I…

(Susan Sparcet): Can you…

Man: …(unintelligible) that.

(Susan Sparcet): …pull up the competing 5 paragraph.

Woman: Okay.

(Susan Sparcet): And we can keep - I think we can keep that up. Because the other thing that we wanted to say is here. So the way that we have you do this is we go through and evaluate the measures.

We want you to look at each of the measures. And each of the criteria in turn and basically hold them up against each other. You know, are they - see which one is stronger.

We are ultimately going to ask you - we're not going to ask you to vote on each of the criteria. But we're going to ask you to vote on an overall decision after you've discussed everything. You compare them on evidence, you compare them on opportunities of improvement, compare them on reliability, etcetera.

We can keep this up and kind of scroll through it as we go. And then NTS also has just some basic -- keep in mind as we discuss -- we want measures to be endorsed but have the broadest application. And then we also want to minimize provider burden.

So those are things that we would you to keep in mind after having this conversation. And we also have invited each of the developers to give a few, very brief opening remarks before we begin the Committee's conversation. So unless there's any questions about process or NTS criteria we can pause, let the developers speak briefly, and then turn it over to the Committee for discussion.

Woman: Right.

Man: (Unintelligible).

Woman: (Mike).

Man: And we're doing this because they are not harmonizable. Okay.

Woman: Yes.

Man: Thank you.

Man: So is there - are there any NQS criteria for determining what's in class? Are there standards, you know, that we can apply to the measures to try to make judgment or is it part of conversation.

(Susan Sparcet): Yes. So actually could you pull-up criteria number 5. Is this it? I can't tell.

This is a table. And so (Susan) actually went over some of it. And it's challenging because there - it's, you know, subjectivity.

The degree in which feel one measure over the other meets a set of criteria within the competing measures criteria. And hopefully we can pull it up so you can see. Because I know it's a little abstractive when talking about it without looking at it.

Just one moment we're getting that slide up. So I think what the criteria says that we are asking you to look at the measures and see if one is better. You know, and that might be one had more rigorous testing or a better results.

It might be that one is more feasible than the other. It just is, you know, that's where we land. And I think we can - if it would be helpful we can go through and give you the votes. As we go if you have any specific questions or - we can kind of…

Woman: And maybe - I'm sorry to interrupt. But maybe what we need to do too is maybe have a side-by-side. We can create it.

That will take a little more time. Maybe 10 to 15 minutes. And maybe we can put the votes against each other. That might be helpful.

But we can see the competing measures. Would that work?

(Susan Sparcet): We can work on that while the developers are presenting? So other questions on the process.

Woman: (Brenda).

(Brenda Lee): My question really is at the end of the day, once we do the comparison is it possible that one of the measures might be kicked out? Or - perfect.

Man: (Unintelligible).

(Susan Sparcet): So the decision will be the best in class is you are voting. And if one is best in class and the other is not, then not will become not recommended.

(Brenda Lee): Okay.

Woman: (Unintelligible).

Man: So even if - now if these measures now have been approved, they've been endorsed. Now they're up in maintenance. But - right, so even though that occurred, we are now going to choose - the losing measure are they no longer endorsed?

(Susan Sparcet): They go for - they're - right now they're - well they're currently endorsed. But they're recommended for continues investment. So the status would be

the measure that is best in class would go forward as recommended for continued endorsement.

And the measure that is not best in class would go forward as not recommended for continued endorsement. The final criteria that we discuss after everything else is competing measures. And it rarely comes up.

So we don't usually go through it. And that's why it seems here that it's overruling.

Man: So what are the implications of the 50% of the folks that you get has - would continue using that measure if it loses it's (unintelligible).

(Susan Sparcet): That, you know, that would be up to the folks that are using the measure to decide whether or not they are going to continue using it.

Woman: And if we didn't endorse (unintelligible).

(Susan Sparcet): That would be up to CMS to decide whether they're going to continue to use it. It's…

Man: (Unintelligible).

(Susan Sparcet): …the conversation is about the merits of the measures. Not about what our decision - the decision made here today says for the future if that makes sense. We want you to look at the two measures and say is one better than the other. And if so which one.

Man: (Susan) can I put a spin on this? I think the purpose of NQF endorsement is to create a pathway for both. Public reporting and accountability visa via including payments related to performance.

And the unfortunate thing about that is we lose the importance of quality improvement as an approach. So I think that's what we're voting on. I think. I mean I understand you're go up the same endorsement.

But I think my perspective that's really what we're doing. I mean they're plenty of measures out there that aren't NQF endorsed that are used for quality improvement and things like that. So I'm just saying I think that's what the issue is.

Woman: I also wonder is there an opportunity to get industry expertise and feedback? Or has that been done between the two measures?

Woman: The measures have been out for comment. You know, (unintelligible) for the pre-meeting comment. But we'll fill out the comment again after this discussion.

We will seek additional comments. Okay. I don't see any other questions around the table.

So I think we will go into the developer presentations before we start. And I think just going in numerical order we can start with the UDSMR. And then - so five minutes for UDSMR and five minutes for CMS. And then we will turn it over to the committee.

Woman: Thank you. And thanks Paulette. (Unintelligible) Dr. (Nesack) are you still there?

Dr. (Nesack): Hello can you hear me?

Woman: Great. Thanks. I'll give some introductory remarks. And we may need your technical expertise because I come at this form a clinical direction.

Let me say first of all I - it feels odd to me that even - we even got here with two competing measures. Remembering back in history I actually served on first technical expert panel when the so called tier 2 predecessor of (GGI) was developed. And a little surprise when I came into that that a good bit of work had already been done.

My feeling is - my opinion is that we maybe would never had gone down this road if there had not been confusion over the proprietariness if that's the word, of the UBS then measures. For some reason there was some belief that they were proprietary or not able to be in the public domain or whatever. I hope from what you've heard today that steered in your mind.

Because the documents in time to (unintelligible) they're having (groupatuities), the right to use these measures were probably metric. But no charge to any user. So in your (unintelligible) hopefully that's laid to rest.

I think it was a matter of concern and maybe was the basis on which the self (unintelligible) needed the new measure built. An odd quirk of this is the first time I guess anybody's ever in a situation with rarely competing measures. Competing measures where PEF has a huge - a single lateral authority to remove the recording of one lumped in and (unintelligible) this measure on the (unintelligible) as of October 1, 2019. Which pointed out I don't anybody to be confused that that, you know, set of the (unintelligible).

Just because, you know, that's not going to be required in that environment. Because I think you've heard today it's used all over this country in more than 7000 published article. Hospitals will continue to use it for various quality reason.

And so hope that's not confusing. So let me just turn to what I think may be deciding factors as you have to look at competing measures. Interestingly the Impact Act, which a lot of things are kind of attributed to.

We had to (unintelligible) because of the Impact Act. The Impact Act promotes particular things that particularly, one is an ability to compare conditions for industry settings. So the two measures that you have before you from UBS 2286 and 2231 are cleared by NTF already to be used in in-patient rehab hospitals, skill nursing facilities, and long-term (unintelligible) hospitals.

So they're actually superior in complying with the purpose of the Impact Act. Another aspect of an Impact Act actually mentions cognitive consideration should be included in functional measures. And the stuff here measure that have you before you from UBS includes the cognitive measures.

Whereas the comparable one from CMS does not. Our statisticians and (Paula) I think can confirm that it's determined that the expression element of cognition actually was the most powerful element within that tool of predicting overall improved function. And it stems to make sense.

I mean we all know that you may have some motor ability but if you cognition doesn't match with that you may not be able to do that. As a clinician looking a quality measure I want to know are we asking the right questions. And are

we asking the questions in a way that allows us to get data that can grab some action from quality improvement, accountability, whatever.

So there several elements that are measures that I think are particularly strong. None of the sensitivity of the scale that the UBS developed measures you. It's a 7 point scale instead of a 6 point scale.

Now someone could argue, it's been brought up today, like, well maybe a 6 point scale is simpler. So perhaps a housekeeper could use it, you know, more easily in making an observation. Or perhaps a family member at home might understand it more easily.

But when we're talking about things that are going to drive quality intervention, changes in process of healthcare delivery, and payment systems, simple isn't always the answer. A 6 point scale blurs some differences. You may not notice the difference between a person who uses a chain and doesn't need to use an assisted device for mobility in a 6 point scale.

So I would say that the sensitivity of scale is very important. And when you get down to a 6 point scale you blur some distinctions that clinically important. Let me address quickly the issue of missing data.

Because I talked to Dr. (Gauge) for a moment in the break to determine that I think she and I might be using missing in a slightly different framework. So I want to see if I can explain carefully. My contention that the UDS developed measures are superior to the CMS measures regarding collecting useful information.

So if you fail to make an observation of a certain element of the clinical test, you know, you didn't observe the person correlating with you. And so you can't score it because you didn't see it happen. A zero can be entered.

And you just would enter zero, like, don't know. Didn't happen, didn't report. So that's at admission.

At discharge it has to be scored. But admission, one innovative thing, I'll CMS credit for, they said well if we're going to give a zero let's tell why we gave the zero. Was it because the patient was too sick or because it was too dangerous to attempt that thing or whatever.

So they've created I think three codes for, like, why we didn't - they didn't make the observation. So when she talks about missing data, she would, I believe say that there's not missing data if they coded it 07 meaning it was - the activity was not done and was not observed for a particular reason. I would still call that missing data.

In others you still don't know at what level the patient performed when they were admitted. And so those - the (Kneader) system that re-coded to a one which is the lowest, most dependent level. And then get scored against how well the person's doing at discharge.

So to the extent that's accurate or inaccurate of how they would have performed if you'd see them perform it. Either you're getting an accurate change in function or an exaggerated change in function. On all the elements in the two UDS measures (unintelligible) put together mobility, as well as self-care for the moment.

The range of what I would call missing data is from - most of them is 0% as high as 2%. The range on the corresponding component element in the competing measures go from a low of 4.7% to a high of 11%. So that's a lot of cases that we may possibly be misinterpreting the change.

There's another difference in how they're structured. Well worse performance versus usual performance. So under the UDS developed measures we asked Professors to score the worst performance in a certain window of time.

And corresponding instructions for - from CMS are score the most usual performance. It doesn't specify the length of time. So first of all it's a little confusing for scores to know what's most usual (unintelligible) class if they one time and most usual.

But imagine a person who goes to the toilet to urinate 5 times in a 24 hour period. Three of those times the person needed no physical assistance to do it safely. Two of those time, maybe random times, maybe at night when they're tired, they needed physical assistance from somebody to prevent a call.

Under the CMS competing measure, most usual 3 out of 5 that person needs no assistance. The UDS measure is designed for clinicians and planning and intervention. So you need to know what a person's worse performance.

That's what you have to staff for. That's what you have to prepare the family to understand. That really what's key to their safety.

So I think that's important. And then broad applicability is I think adventurous here. There's been a lot of questions about how can we unify, modify, and, you know, make things more holistic.

The UDS measures go down to age 18 instead of 21. It may not be a huge thing because the rehab population is heavily skewed towards older people. But diagnosis are skewed by age.

So the 18 to 21 population they're more likely to get traumatic brain injuries, traumatic spinal cord injuries, multiple trauma for motor vehicle accidents for example. I think it's important for quality measures to be equally concerned about that whole range of people. The broad applicability it's used in the VA system.

It's used throughout this county. It's in 50 other countries around the world. The basis of most published literature in the rehab field is based on the elements of these two UDS sponsored measures.

And so these are some of the keys that I would say in having to make a judgment about superiority I would say that these weigh in favor. Reading through…

Woman: (Unintelligible).

(DexAnne Aquila): …okay.

Woman: I was just going to say. I thought that that was your conclusion.

(DexAnne Aquila): Conclusion. I took a breath. The ask Paulette is there any statistical at stake or something that I left out?

Or that you should say quickly? And then I will be quite and turn the podium over to our Chair.

Paulette Niewczyk: You - thank you (unintelligible) that was excellent. I just wanted to add just maybe two points. In addition to the age elements it's not only there's, you know, that we - that the UDS measure goes down to 18 years of age.

That is important. But it's for all payers. So the CMS measures are intended for Medicare beneficiaries.

And for the most part that includes people ages 65 and older. But that's not the entire patient population that post-acute care will treat. You have workers that are injured on the job.

You have, you know, Veterans as (DexAnne) mentioned. You have young adults that have private insured - insurance companies that are reimbursement. You have those that can self-pay.

You also have those that are not yet 65 and eligible for Medicare, yet may have a stroke and need additional post-acute care. And don't have any health insurance. So when we talk about disparities that is a segment of the population.

And because you're not getting payment reimbursement these measures would in essence not include those patient populations. So when NQS talks about broad application, our measures are superior based on they are broadly applicable for the entire post-acute care population. And they do minimize burden. We have….

Woman: Paulette…

Man: You're finished.

(DexAnne Aquila): …okay. Thank you Paulette.

Woman: Thank you Paulette.

(DexAnne Aquila): I'm sorry I used more of your time. But thank you madam chair.

Woman: All right. CMS I'll give you a little longer time.

Alan Levitt: Okay. Thank you. I'll start off and I'll turn it over to (Anne). I came from an (unintelligible) that was one of the 30% that you've seen almost from the beginning.

In fact we were - we moved from (Barthall). And (Barthall) was actually a therapist on my unit. Actually retired before I ran the general Rehab Unit at Montebello at first.

And I probably one of the most avid users of the custom reports of UDS up until I had to retire or I had to leave the Board in order to take the job at CMS here in 2013. So we do not as an agency - and I'm not speaking just for - we don't view these as competing measures. We really don't.

I know the NQF does but we don't. We developed our measures based on the mandate that the Impact Act gave which was measures within these different domains but it was based on standardized assessment data. And the data we're talking about was developed as part of the Pact PRT.

It was developed for multiple purposes. It was developed for longitudel care of patients. It was developed for quality measures.

It was developed for resource use. And that is the data that is being used now in these measures. That again are being used in not just the two measures we're talking about but in the two measures we will be talking about also uses the same data.

Oh actually (unintelligible). The LTACT measures which was also mandated by a different statute. The Bi-Partisan Budget Act of 2013 again uses this data as well.

This measure is also now adopted into the Smith Quality Reporting Program. Which again is 15000 Smith's. That data's also now being collected in home health as well.

Certainly the hope and the long-term would be to have measures within all post-acute care settings if not all healthcare settings based on standardized assessment data. And I am - this is what we're talking about. If IRF and hopefully IRF will and we were using the same before the PPS came out.

And we would hope that IRF who continued to value that data and those reports will continue to do that. We, you know, we don't want to stop that. We don't want to stop that in Smith's.

Smith's collecting data. You know, if we're collecting, we're using function (jada) with the GG if they have other items. Continue to do that.

Same thing in home health, same thing in Health Act. I mean essentially each setting could continue to do what they've been doing but for the standpoint of looking at a cross-care in terms of items, in terms of the measures, this is why we do what we've done. Just before I turn it to (Anne) quickly to clarify, again these - this wasn't an issue of proprietary or not.

This is an issue of items that could meet the different goals that were first established within the Pact PRD and the objective there. That continue to be the objectives that came across in the Impact Act. Other things the thin items it wasn't a unilateral removal of a measure that we're talking about here.

What did occur is as in all rule making, the decision was made that the payment system within our group could be better modeled based on the GGIM's versus the Stem items. And that was proposed to the general public. Public comments were received and then it was finalized in that way.

There have been no measures related to any of those items that were being used for payment. (Anne) do you want to take specific, I think.

(Anne): Sure. Just take a moment to talk just in terms of the data elements. So the data elements that have been a main topic I think with discussion.

So the data elements that are used in 2633 and 2634 were developed by an inner disciplinary team starting in 2006 as part of the post-=acute care reform demonstration as Alan mentioned. And basically that group used all the existing knowledge and all the information from existing measures, like, (unintelligible) and the information on other data sets. And the literature to try and develop a set of items that would actually work across all post-acute care settings.

And so there were a lot of discussions with industry experts across ERF, SNIP, EL PAC, and home health agencies. And so there were some refinements in terms of thinking about how to define items because of less informant. So just a quick example for eating, the SIM allows if somebody is

getting fed through tube feedings and not eating by mouth, you know, that was an issue that we thought was causing some misfits.

So measurement issues with the SIM items. And so we actually split up eating by mouth, and getting fed through tube feeding. And so we actually had two data elements on the original (PAC) pier D data set to address that.

And we found actually that the tube feeding didn't work. But of course the eating item actually did work well. So we do see some activity not attempted code because somebody doesn't eat by mouth.

And so it's, you know, really different perhaps to a particular person instead of administrating a tube feeding. Which is more of a kind of medical procedure. And so there's things that we learned from the existing items on all the data sets that we tried to make as simple as possible to make sure the data was basically measuring one construct as cleanly as possible.

In terms of the issue about cognitive function. I do want to clarify that, you know, these measures are self-care, mobility. I would think CMS would like to measure - to have something related to cognitive function in the future.

It's required by Impact Act. But that's not work that hasn't been completed at this point. They do have some work going on in that area.

There was an article actually in 1994 in Archives of Physical Medicine Rehabilitation about the (Stim) that basically said that cognitive function and motor function are actually very different constructs. And that they - you shouldn't - when you're looking at data across diagnosis you shouldn't merge those data together. And so that's been something that's been out there for a while.

And so as I think I said as part of our work we also did analysis looking at self-care mobility and found that those actually split out when you looked at some of the patients across all the settings. And so Dr. (Clohan) mentioned that cognitive function is a predictor of outcomes. And actually it is a risk adjustor in our functional outcome measures for both self-care and mobility.

So we do have cognitive function as well as motor function as risk adjustors. Yes. Okay.

Alan Levitt: Okay.

(Anne): All right. I guess that's it.

Woman: Thank you both. (Susan). Well - right here (unintelligible). Let's…

(Susan Sparcet): We're pulling up the table of both. We're putting that together.

Woman: Oh would it be helpful to go back to the decision treat here.

(Susan Sparcet): Yes. We can pull that up too. Decision tree.

Woman: And maybe just as a group go through.

Man: Okay.

Woman: Yes. Now…

Man: Just to make sure that we all agree with her.

Woman: (Unintelligible) tree. I have it in front of me.

Man: Oh yes.

Woman: All right, you can't see it but I think we can click right through. Does the measure meet all core criteria, making it suitable - yes. Are they potentially related and yes.

Compare the specks. Same concepts for the measure focused or the same target patient population as another endorsed or new measure. Yes or no.

Any - that's kind of where I stopped first. Because it seems to me that patient population is one of the points of difference.

Man: Yes.

Woman: Am I right?

Man: Yes.

Woman: Okay.

Man: Okay.

Woman: Then if - excuse me. Can one be modified to expand the population when we were told potentially you could someday? It would be in response to (Anne) question it would require redoing the measure and bringing it back for us?

Correct? Okay.

Man: Okay.

Woman: Now I am - here is where I need help.

Man: Right. Do we just have question or…

(Susan Sparcet): Yes.

Man: Okay.

(Susan Sparcet): As soon as she's done.

Man: Great.

(Susan Sparcet): (Unintelligible).

Man: Okay. Let's ask for opinion. (Peter) you were first.

(Peter Thomas): I don't know if you're going to like what I have to say. But I'm going to be a non-compliant patient for a moment. And I'm not really sure why we're doing this.

I got to be honest with you. I - if there's a question about winners and losers in terms of the - these two measures, the CMS RTI measures have already won. The regulatory process is phasing out.

The (Stim) for the IRF system, the measures apply across all four settings of post-acute care. CMS is clearly moving forward with that. There are statutory requirements that are - have been put that in place that kept that in motion.

There's a competing set of measures that seem to make some sense in terms of the population being broader than Medicare. And therefore useful to non-Medicare beneficiaries. The CMS measure developed for himself values the reports for UDS has put out to its recipients who value those reports and say at least of them will continue to rely on them.

Even acknowledges - even suggest - not suggests, stated explicitly that these are not competing. So I don't really know why we're being in a sense forced to make a what I view as kind of an artificial decision. These measures will continue.

RTI CMS will continue to use the measures it uses for the post-acute care unified payment system. As that gets developed in the future and to the extent that they're already being required to regulation to use what they're using. They'll continue to go about doing that.

If we by some - for some reason chose to side with UDS as opposed to RTI, you know CMS is going forward with those measures anyway, without NQS endorsement. So I don't understand why we're kind of being put into - in a sense into this position. I just - it doesn't make sense to me.

Man: (Alisa) yes.

(Alisa): Let me answer that. It's a great question. You know, we, you know, as I mentioned before we're not just trying to get the measures that are scientifically sound out there. We also trying to make sure that we're addressing some of the concerns we've heard lately about measurement.

You know, we used to be very proud to say we have 600 plus measures. And then we started hearing about the burden of measurement. And so that is really important for us to make sure that we have the right measures that are out there.

That we are eliminating duplication where we can eliminate it. And if we - you're answering all of these questions yes, it appears to us for our definition that these are very much alike. At the least still probably about maybe the same.

So we want you to make the best decision that you can. We want to make sure that beyond looking at the specifications we're looking at implementation, we're looking the burden it can cause at all levels of the healthcare system. So that's not as apparent when you're looking, you're evaluating the measures.

They're individual merits or as you may be looking at them side by side, we're thinking about it and the effects really to the health care system abruptly. If you decide, like you did 3 years ago, as this Committee did, ultimately that decision you're making a recommendation to the CFAC. And the CFAC has stewardship over our entire portfolio of now 400 and - no 540 measures.

We have less measures in there. They may decide that it is worth having these measures endorsed as well. But they need the interest from you as the experts in the field.

So we may go down the same path. But we want you to have that discussion. Thinking about the specifications, thinking about the scientific merits, but also

thinking about implementation. And what this may mean to have these 4 measures out there.

Man: Okay.

Woman: (Linda).

(Linda): Yes the (Stim) metrics have been around for a long time. And are proven metrics. The care model are newer metrics.

And they're still under development. My concern is if we pick one and not the other that we might win at competition. You know, if we have another measure that's coming up through the ranks, it's new, it doesn't have an opportunity to test it before it gets here

So I just - I would agree. I don't see these necessarily as competing, yet. I yes - that's it.

Woman: (Gary).

(Gary O'Mally): Yes and I agree with both (Linda) and (Peter's) comments. In a sense these measures are used in different realms with different purposes. And in that sense they really are not competing.

So I think it's a miss-number. I think the healthcare system would get along quite well with both measures in tact doing what they're doing, in their different settings. Each with a different purpose in mind.

So no one's going to argue about the value in the measure and rehab scores. And rehab improvement. No one's going to argue with the Impact Act face

measures being the basis for the entire healthcare system across all sites in post-acute care.

And that's the way it is. So I would - I kind of agree we're - have a mutiny on - but (unintelligible), you know, let them both go. They're not getting in each other's way. And they're both adding something to the healthcare system.

Woman: (Don).

(Don Casey): Well to me it's more of a fundamental philosophic question, because I can say this back in the day, we were trying to cram as many measures into the pipeline. Knowing that most of them weren't' even used for payment and public reporting. Which are - is my simplistic translation of accountability to be honest.

So now we're sort of going back the other way. Which - I understand that. It seems as though the fundamental issue here is more foundational to the importance of NQF visa via a process that I think is, you know, it always needs work. But it's a valid process.

It's a consensus development process. And it is -endorsement to me means something more than just you can use it in the payment system. It means that it's got validity and reliability that's been testing by experts through a giant ringer.

And I don't know. It seems like maybe at some point in time we ought to portion off measures rather than get them going. And maybe that's what we're, you know, get rid of - getting rid of them. Maybe that's what we're doing here.

It just seems as though there's so much scientific evidence that has generated around the (THIM) score as everyone has said. And I, you know, even I have some question about utility and public reporting and accountability as we know it. So I just think that I'm on board with this idea of not making an either or decision here because - and I know that's a larger scale discussion in NQF. But I'm on the side of saying let's keep both and let's go forward.

Man: Yes. I would love to get a better understanding of what was going inside (CSAC's) collective head to help us understand better why this sort of strong recommendation to have this discussion was. You know, there's burden, there's, you know, need for parsimony where that's possible. Was there more than that?

Woman: You mean for that (CSAC) going forward? Because this would be the first time that they're making that decision. The Board made it prior to…

Man: Okay.

Woman: …you know, 2 years ago.

Man: I just would love to understand what was going on in their discussions that might help us.

Woman: Yes a lot of it was around the burden. A lot of it was around and I think it goes back to (Don's) comments about, you know, that's at the height of hearing about measures that matter. And recognizing that not everything that even we have endorsed really has been moving the needle in terms of improvement.

And so I think they were trying form their vantage points to make sure that we are truly not flooding the field with just measures that have gone through a consensus process. But we can say that if we're saying they're the same target population being measure or almost the same measure, let us make them - let's help the field set and say…

Man: Okay.

Woman: …we - this is the best one.

Man: So if we say something like we really feel strongly that both measures matter and there's good evidence for that. Then…

Woman: Yes.

Man: …maybe there's a rationale to keep them.

Woman: I think you should have a very strong statement coming from the Committee if you feel so.

Man: Right.

Woman: It sounds like the consensus.

Man: We might get in trouble.

Woman: And because what could happen the (CPAC) has rationale for overturning the standing Committee's recommendations. But it's really based on 3 things. We don't want the (CPAC) re-adjudicating standing Committee's decisions.

It's whether the process wasn't followed, the criteria weren't followed, or the measures don't add value to our portfolio. But they don't - what they may do is kick it back to you, quite honestly. And so we may go through this ping-pong.

So it would be good for you to have a very strong statement. It sounds also like you have concern whether or not they're purely competing. In terms of the target population that's something that I just stuck on from the current discussions, so.

Woman: I was on the (CPAC) if you recall.

Man: Yes.

Woman: And I also was a co-Chair of this Committee. And was one of those individuals therefore who has explained the Committees discussion to (CSAC).

Man: Are we getting you trouble?

Woman: You are not getting me into any trouble at all because the situation was you're measures were grant (unintelligible). You were still sending us testing I think part way through our deliberative process. What we now are - we are now the situation which is exactly what the process is supposed to do.

We've had - base has been implemented, they've been running for 3 years. We are now taking a fresh look. And I think that whatever decision we come up with that is precisely the way we set stage for people.

Woman: I have a question for (unintelligible). So my question is just to make sure in my non-clinical head that I'm really understanding it. We have these two competing measures.

What we're saying is they may or may not measure similar - exactly similar things. But the key in terms of trying to reduce measurements isn't who is the end user of the measure. Because these measures are being used differently at the end user.

But that the burden is reduced on the clinicians who have to fill these things out, right. So currently because there's even - there's different end users but there's two measures. So that 1150 (ERFT's) or whatever out there, whoever is filling these out, they still have to fill out each of these measures, right.

So the concept - and I'm correct in thinking the concept here is to reduce the burden on the people who are filling them out. Not to have less out in the field. But are you saying that whether we vote to reduce the burden or not they're still both going to be out there because of the different end users.

It's a little complex. Somebody was talking about cognitive issues with all of these measures running around in our brains. But is - am I getting it at that correctly.

Woman: Yes.

Man: Yes. I'll take a crack at that baseline. And it may help for us to define what we mean by burden as interests in discussing of this is exactly what you just pointed towards.

So we say a measure is competing. If you can imagine a scenario when the measured entity would be simultaneously responsible for coding or documenting essentially the same thing to fulfill, you know, a patient who is both the numerator and the denominator of two measures. So for example, just to make it a little more concrete.

If one measure was used for payment and another for accreditation. And you're responsible to report on both, they have essentially the same elements. Or some areas of overlap where it's caused for concern that's additional burden, that's what we would consider competing measures.

Woman: And (Peter) go ahead and then (Brenda).

(Peter Thomas): I would just say two things I've heard in response essentially are whether these are competing are not and I really do believe that, you know, Medicare is a big payer. But it's one payer. And they're a lot of other payers.

And, you know, there may be reasons why a rehabilitation hospital would want to continue recording data under both measure sets. To satisfy certain state requirements does that fit the benchmark against previous performance to not have to retrain their entire staff on how to do, you know, a quality metric (unintelligible). Even though ultimately their going to have to do that because that's where CMS has moved.

So the reg's required that IRS record data under both systems. And now they've eliminated that requirement. And they - their reporting data under the CMS measures - or will be shortly.

If a rehab hospital wants to occur the burden of continuing use that UDS base measure set. Why would we get in the way of that? It's going to die on the vine because entities - IRS doesn't feel that there's value in that.

Then so be it, that's not our choice, that's not our role. But both of these measures independently were endorsed by NQF as recently as today. As in terms of our recommendation of these.

It just strikes me that there's really no reason to go further than that. And I would be more than happy to try to work with anyone else to try to put together a statement to that effect.

(Brenda Lee): So I - (unintelligible) speaks with me in a big way, right. Like what - essentially let the markets work it out. But I do have one concern.

So let me ask as a question for - confirm if I'm understanding correctly. If an IRF uses both the (Stim) instrument and the care tools so that they would be able to calculate both these measures. Does that put added burden on the patient?

So for example would the patient have to go through additional assessments? Because that's a concern, right. Like, I'm already tired.

I'm working hard. I'm going through PT. And some of this may be painful or, you know, difficult and now I got to climb the stairs twice. So it's a question. I don't - is it…

Woman: No.

(Brenda Lee): …add a burden for the person scoring?

Woman: No.

(Brenda Lee): Or does it add a burden as…

Woman: The person…

(Brenda Lee): …doing additional…

Woman: It's the person scoring not the patient.

(Brenda Lee): Okay. Yes. Thank you.

Woman: (Unintelligible) for more confidence but - so but from the scoring perspective you don't sit the patient during these different testing.

Woman: (Beth).

(Beth): Just a couple of questions. There was a (Stim) tool that needs a - it's a theory of a loss of impact. So that's why there was another one that was developed.

That's the - our first question when the impact that was passed and then the measures were developed. And then the second question would be do the developers ever have a (unintelligible) together over time or does it stay competing?

Man: The GG items we're talking about were specifically developed for a purpose of being able to be used in multiple post-acute care settings at the same time. And again for quality, for longitudel care, for resource use. The (Stim) items

were, you know, and have been successfully used in the IRF setting, you know, again for IRF patients successfully.

Woman: Let me just - you're not limited to that you can take the - I think the measures that have already been improved by NQF as UDF measures are precisely identical measures in all sense of the setting. The setting that you're CMS measures are being used in, I think you mentioned Dr. (Gauge) that they've now gone out to (unintelligible) sites and (El Pac) sites. I don't think the numerators and denominators are precisely the same in all three sites.

I could be wrong about that matter. But I don't think that it's the identical measure in all the three sites. Correct me. I may be wrong.

Woman: We're going to…

Man: Yes. I don't have the counterpoints (unintelligible)…

Woman: …be closed.

Man: Okay.

Woman: And we do need to have a vote in order to conclude this issue. And I am turning to my colleagues here as to what we are voting on.

Woman: Yes. So (Susan) is going to cue that for us.

Man: Right.

(Susan Sparcet): You're just having a little side bar about how you're going to vote.

Man: Okay.

(Susan Sparcet): Okay. I think what we're going to do is offer you three choices. You can pick the UDSMR measure, you can pick the CMS measures, or you can pick - I think - I want to recommend both.

So it's one of the measures or both. And then in terms of getting to consensus we would say this is where we're a little fuzzy. Whoever - if there's, like, not 60% of folks voting I want to keep both then we would or if more than - then we would go with you'd have to pick one.

Woman: (Unintelligible).

(Susan Sparcet): Yes. So now we're going to pull up that vote.

Man: Quick question. There was - I think maybe an effort to try to pull up the vote side by side for the two measures. Did that…

(Susan Sparcet): We now have that yet.

Man: Okay.

Woman: (Unintelligible).

(Susan Sparcet): So give us a moment to get that...

Woman: One more clarifying question regarding the two measures, so at IRF's they are in most cases reporting on both of these measures. So the staff or clinicians are doing both of these. Is that an accurate statement?

Woman: Right (unintelligible) left. Yes (unintelligible).

Man: Right. The items…

Woman: (Unintelligible)…

Man: …for both of the measures.

Woman: …acquired it switches over.

Man: Right. So once again the items. So the items would be data elements that are the components measured are both being collected currently.

Man: Great.

(Susan Sparcet): And here are the side-by-sides so you can - can you make it a little bigger. You can see fairly close. Would folks like me to read this out or would you rather look at it.

Man: I think everybody's gathered around.

(Susan Sparcet): Okay.

Man: We're okay. Look at…

(Susan Sparcet): The measure on the left is the UDSMR measure and the measure on the right is the CMS measure.

Woman: All right.

Woman: We are but you…

(Susan Sparcet): (Unintelligible).

Man: You're vote.

Woman: Right.

Man: Yes.

Woman: And for picks…

Man: That's up to you.

(Susan Sparcet): Those are your votes. So the liability and validity your votes to agree with the method panel recommendation. And both passed method panel, so.

Woman: Well thank you anyways.

((Crosstalk))

(Susan Sparcet): All right.

Woman: (unintelligible) CSM,

Man: Give them all three.

(Susan Sparcet): And we're going to the first set and then the second set. So we're going to do two votes. One on each set of measures.

Man: Competing measures best in class NQF 2286 versus NQF 2633 is now open. You're choices are A 2286, B 2633, or C recommend both measures for maintenance and endorsement. Should A and B receive consensus, we will then vote for those individually without including option C.

(Susan Sparcet): Okay so we have 17 people voting.

Man: And my computer - my delayed sites. So I have to give you verbally. Maybe course C.

(Susan Sparcet): Okay. So we're - can you enter a vote for him.

Man: (Unintelligible).

Man: Voting is now closed for competing measures best in class.

(Susan Sparcet): Then we're going to add that into this account.

Man: Oh sure.

(Susan Sparcet): Okay. We're good.

Man: Option A received 2 votes. Option B received 2 votes. Option C recommend both for maintenance and endorsement received 13 votes.

(Susan Sparcet): So that's 13 to 4. And so - can you pull up the percents. But I believe we are into a consensus. And…

Man: That is consensus.

(Susan Sparcet): The committee has decided that we are looking at recommending both at 76%. And now we'll do the same vote for the other measures. And again we have those votes here.

UDSMR on the left, CMS on the right. Apologies we don't have counts for usability on one of them.

Man: Competing measure best in class NQF 2321 versus 2634 as well as recommend both for maintenance and endorsement is now open for voting. Option A NQF 2321, option B NQF 2634, or option C recommend both for maintenance and endorsement. Voting is now closed.

Option A NQF 2321 2 votes, option B NQF 2634 2 votes, and option C recommend both for maintenance of endorsement 13 votes. Recommend both for maintenance of endorsement has reached consensus at 76%.

(Susan Sparcet): Okay. Thank you. SO we will take (Peter) on his offer to write that statement and share it with the Committee.

Woman: And I also I would add if I can ask RTI and CMS and UDSMR it sounds like that there was some agreement from you both on perhaps maybe the target population not being quite the same. And there might be some complementary aspects of having the sets of measures exist. It might help for you to work together to write a statement for the (CSTAC).

Or maybe comment during the public comment period. Just some words of advice in this new territory that we're in now, so…

Man: So I think we're gotten towards the end of our time together. And I just wanted to first thank our measure developers for handling what is

undoubtedly a stressful situation with a lot of grace. So well done to the two teams.

Thank you very much for that. And big thanks for our Committee for all the efforts for this afternoon. Really appreciate your time. I'll hand it over to you (unintelligible) Chairs for closing remarks.

Man: Need to do public comment.

(Susan Sparcet): So we'll now open the lines for public comment. If there is anybody on the phone who wishes to speak or submit a comment via the chat or anyone in the room, please do so now. And we'll give that a moment people to submit or speak. Okay hearing no comments we'll go to next steps.

Man: Yes. All right upcoming next week we have the post-measure evaluation web meeting on June 25, Tuesday 2 to 4 pm EST. As well as on Friday June 28 2 to 4 pm Eastern the post measure evaluation web meeting number 2.

Potentially if necessary we will schedule a third post measure evaluation web meeting based on the doodle poll that most of you filled out. Looking at either Monday July 1 2 to 4 or Tuesday July 2 1 to 3, both pm Eastern. And then other key dates that will be upcoming will be the report to be posted for public comment on August 1st for 30 day public commenting period.

As well as the draft report post comment call on September 25 from 1 to 3 pm Eastern. Our project contact information. And any questions.

Man: (Unintelligible).

Man: Yes. Sure.

(Susan Sparcet): Just to add we'll discuss the reaming three CMS measures and then all of the (CAPS) measures on the webinars next week. We'll share an agenda by COB tomorrow.

Man: Could I just - I know we're in the air about the 28th, but if we could do home care (CAPS) on the 25 that'd be good because I'm not going to be able to make the 28.

Woman: You can (unintelligible).

Man: (Unintelligible).

Man: Okay.

Man: (Unintelligible).

Woman: I want to thank everybody for a very difficult day. And many of you probably don't realize quite how hard (Susan) and (Joshua) - (unintelligible) (Jordan), and (Sam) and (Alisa) have worked with us. (Chris) and I have been on more than (unintelligible) call around this set of measures.

(Chris): Times 4.

Woman: And they have - they've gone way beyond churning out very good material. And so until we talk again on Tuesday.

(Don Casey): Well (Chris) and (Lee) also great job on your part as Chairs as well, so.

(Chris): Thank you (Don). Thanks - yes. Great job everybody. I think everybody was engaged. And we actually did a lot of work. So thank you.

Woman: Good work.

(Susan Sparcet): Safe travels. Meeting adjourned.

((Crosstalk))

Woman: Oh boy thank you.

END