Male Operator: Welcome and thank you for standing by. Today's conference is being recorded. If you have any objections, you may disconnect at this time. All participants are in a listen-only mode until the Q and A section of today's conference. At that time, you may press 1,4 on your phone to ask a question. This presentation is not intended for the press and the remarks are not considered on the record. If you're a member of the press, you may listen in, but please refrain from asking questions during the Q and A portion of the call. If you have any inquiries, please contact CMS at press@cms.hhs.gov. I'll now hand the call off to Maria Edelen from the RAND Corporation.

Maria Edelen: Thank you. Can everyone here me okay in the room and on the phone? Okay. Good. Hi everybody. I'm Maria Edelen. I'm a Senior Psychometrician at the RAND Corporation and the Project Director on the RAND contract with CMS for the development and maintenance of standardized post-acute care patient assessment data. And this work is being conducted in support of the IMPACT Act.

Thank you so much for joining us today. It's nice to have some people in person as well as on the phone. In addition to this group of folks here, we have a lot of people on the phone and we are excited to be sharing with all of you some of the early findings from our national beta test. This went into the field over a year ago, or almost...right around now we were getting into the field and we finished up in September. So, we have a lot of material to cover, but we have deliberately set aside a lot of time for questions because we're here really to give you the opportunity...a little bit of a taste of what we learned and give you an opportunity to react and respond to that.

So, let's get started. For those of you on the phone, I'll just be saying 'next slide' each time. So, this is the agenda which runs from 12:08 PM TO 4:00 PM. We'll start off with a brief overview and some overarching findings from the beta test. Then we'll go into the field test results for each of the data elements in the categories that were tested. And as you can see, as I said earlier, we do have a lot of time set aside for questions in a few places. And the times on this agenda are estimates. We laid it out and we think it's going to be relatively accurate. There may be a little wiggle room in there, but for those of you especially on the phone, we're going to do our best to hit the question periods as indicated on this agenda.

So, hopefully that will make it a little bit more convenient for those of you on the phone. And then just a little more sort of housekeeping, this is a list of several terms and abbreviations that are contained within the presentation. It's really here for your reference. I don't feel like I want to take our time going through all of these, but if you're able to flip back to it...and also, most of these I hope are relatively familiar.

We are also going to be referring to the beta protocols throughout the presentation. For those of you who are in the room we have printed copies. Does anybody not have a copy who needs one? Okay. And for those of you on the phone, the protocol can be downloaded from the CMS IMPACT Act SPADE webpage. If you don't already have these protocols, the URL for that page is listed on a few slides later on in this presentation. And then the specific pages that we are going to be referring to will be shown in the presentation as well. So, next slide.

So, RAND has been conducting this work under the guidance of a CMS team in the Division of Chronic and Post-Acute Care which is being led by Stacy Mandle. I'm going to hand this off to Stacy in a few minutes to introduce the rest of her team and kick things off, but I first just wanted to acknowledge a few of the many colleagues at RAND and elsewhere who contributed to this effort. Emily Chen, whom you will hopefully meet shortly will be involved in some of the Q and A with us, is the Co-Director. Sangeeta Ahluwalia, who is not able to join us today, played a large role in the assessor training and data collection efforts. Anthony Rodriguez is on his way. He's the Senior Psychometrician on the project. And we also have Susan Paddock as the Senior Statistician.

In addition to those folks at RAND and many others at RAND, we had a lot of support from sub-contractors at ABT Associates, which is a team led by Terry Moore. Qualidigm which was under the direction of Ann Spenard. A group of folks from Atlas, and finally some researchers from Northwestern.

So, with that I'm going to hand this off to Stacy for a few slides and she's going to give you a brief overview. Take it away. Thank you.

Stacy Mandl: Let me make sure I know how to work the slides. There we go. Got it. Well, thank you all. Welcome. We're so glad that you joined both in person and on the phone. This has been a lot of work. I

was just reflecting back to September of 2014 when the IMPACT Act passed and the roller skates that we all had to put on, not just CMS and not just our contractors, but all of our stakeholders as well. And we really appreciate this journey that we have been taking together. Looking through some of the materials that I'll touch on today and the amount of work that has gone into not just this body of work, but all of the work with the measures and the standardized assessment data elements has been tremendous. So, we really value your partnership with us. And we're really excited and very grateful to RAND for holding this forum to kind of wrap up the national beta testing.

So, I want to thank some folks also. I definitely want to thank my partners in this work really who lead the work was Dr. Tara McMullen, who was unable to be here. Her father passed away and she is in Reno. She attended his funeral yesterday, so unfortunately, she was not able to make it. But she's extremely excited about the work and very proud of all the hard work that everybody has put in to this. With our contractors but also with the public and the stakeholders as well.

Charlene Van, who is the contract officer or representative was not able to make it. She had an injury, but I believe that she dialed in. She's done a fantastic job meeting the contract oversight in this work. And my partner, Mary Pratt, who's here with us.

I also believe that we have some folks from CMS on the line, so I'm really grateful to them for joining this important work and listening in as well. Maria and the team have done a phenomenal job with APT as well and Qualidyne in this work. Not just the national beta testing, but all of the work has been a huge lift and we're so glad that we are at this point. But most importantly, we want to thank the providers who have been involved in the national testing and spent the time and energy and resources to be a part of this. And overwhelmingly so, our gratitude to the patients and residents that allowed us in to be a part of this testing. Thank you so much to everyone. I'm sure I've missed someone, and I apologize, but the gratitude is there.

So, just kind of a refresher for anybody who may not know, the IMPACT Act was a bipartisan bill that went through and was passed in September and signed into law in October of 2014. That's when we put on our roller skates. And the intention behind the standardized assessment data for post-acute care was to obviously improve quality outcomes and enable for comparability

across providers in post-acute care, to enable information exchange. The law does clarify or specify that the data is to be made standardized and interoperable. So, we've been working with our partners...I knew I'd forget someone. With Office and National Coordinator, Liz Flanahall is here with us. She's been a phenomenal partner with us this whole way along. Enhance care transitions and coordinated care obviously, and person-centered goals that are individually driven.

So, in case you're not aware, the post-acute care providers involved are the home health agencies, the in-patient rehab facilities, the long-term acute-care hospitals, and the skilled nursing facilities.

When we talk about data element standardization, and you're going to hear a lot about it today, what we're talking about are those core standardized assessment data elements that are identical in definition and that are a part of the assessment instruments that are required of those post-acute care providers. And that's what we'll go into. And what was sort of the sweet spot. I love this picture because of not just the colors, but because it's really about having that core information that can be exchanged but can also be used for other purposes.

So, I'm a nurse by background, and to me this body of work is sort of like if you had a hospital that had a different assessment for every kind of unit, and they collected very similar information, but it was never exactly the same say you had to sort of manually input information to have sort of this sense of how somebody was doing across the trajectory of services within that institution, and say, "Well, that doesn't quite work. Why don't we have standardized core assessment data across the unit?" That's what this would be like.

So, what RAND did was specifically test standardized assessment data elements for use across both acute-care in five categories. That's what the law requires, at least these five categories: function, cognitive function, special services treatment and interventions, medical conditions and co-morbidities, impairments, and then other categories. So, RAND tested a set of data elements in some of these categories for us. And all the way along through this process, we applied very specific principles. And these are principles that were important to providers, to clinicians, to stakeholders, to patients and those were the following: the potential for improving quality, of course, is at the top. Improving care

transitions and person-centered care and care planning, improving care practices and patient safety, the ability to use that data for quality comparisons including value-based payment models, to support clinical decision-making and care coordination. Those were some of the very most important principles related to potential for improving quality.

The data had to be reliable and valid. So, a big chunk of testing had to do with ensuring that the data were reliable and valid. And that's what we're going to hear about today. And the feasibility in post-acute care that the data that we would use to satisfy the IMPACT Act be clinically meaningful. That it be relevant to the clinical workflow and that it could be information that could be exchanged and useful in the exchange of that information.

And then, finally, utility for describing the case mix. The ability to use the data to inform payment models and to be able to look at varying severities. That was very important.

Through this whole process, and this was just with the standardized assessment data element work, conducted by RAND, there was a host of stakeholder engagements. And this is so important and many of you I think, actually joined a lot of these. So, there's a lot of work, not just on our end, but on your end. Keeping pace with everything, dialing in to the calls, joining as you are today. And we very much appreciate it.

But there are technical expert panels, there were special open-door forms, blueprints, public comment periods from the rule, and we heard you, small group discussions with PAC associations, and then dialogue through the clinical staff during the testing. That was very, very important to get that sort of boots on the ground input. To get that consensus of experience from those clinicians in the field.

And with that I'm going to hand it over to Maria. Thank you again for joining.

Maria Edelen: Okay. Thanks Stacy. So, I've got two screens going here. You're going to have to bear with me. I'm going to catch this one up. Okay. So, in this next series of slides, I'm going to very briefly describe the data element development process in a nutshell that led to the beta test. And then I'm going to discuss the design of the national beta test and provide a sample description.

So, our overall project goal has been to develop, implement, and maintain standardized post-acute care patient assessment data. The project which started in September 2015, just a year after the roller skates were put on, has been conducted essentially in three phases. The first six to eight months focused primarily on information gathering. So, this is really just getting a sense of what's out there, what's being collected currently in each of the post-acute care settings, where are the synergies in these assessments that we might be able to leverage to help us with standardization, and also what are the gaps?

So, there are these clinical categories, and there's what's being assessed, and then there's some gaps in there. And so, especially in particular for the gaps, we were looking through the literature to try to help fill in some options for how we might assess some of these clinical categories.

So, we conducted extensive literature reviews and we also consulted with clinical and subject matter experts very deeply in this first several months. We convened an initial technical expert panel meeting during this time mostly to determine our priorities and help set the course for our pilot testing.

So, there were two rounds of alpha testing that were conducted starting in August of 2016. And then from there we went through the results and modified some of the data elements. And then tested several other data elements in the second pilot test. All of that ended in mid-summer of July 2017 and we used all of that information to prepare for the National Beta Test. And that's really what everybody wants to hear about. So, I'm trying to get there. The assessment protocols that were tested, as I mentioned earlier, are on this website. And we will also be referring to them throughout the presentation.

So, Stacy showed you the categories that are called out in the IMPACT Act and RAND...we covered the majority of these categories. The functional status was one that another contractor, RTI, was covering. And so, that one we weren't involved in except to be aware of their progress. But we tested data elements in beta. By the time we got to beta, these are the data elements that were included. We tested several data elements to assess aspects of cognitive and mental status. We covered both a pain interview and a staff assessment of pain. The beta test also included several data elements assessing sensory impairments, incontinence as well as a group of nutritional approaches and other services and

treatments.
And finally, the beta testing included data elements from these other clinical categories that were of interest for various reasons: the assessment of care preferences, global health, and medication reconciliation.

So, I'm going to be presenting results from each of these categories in turn. But first I want to just describe the design of the beta test. And for those of you who are on the phone, I've not been saying 'next slide', but we're now on slide 15 which is titled 'Design'.

So, all of our data collectors were trained...they were either research nurses who were recruited by our project team, or facility and agency staff assessors who were essentially nominated by the provider organizations when they signed up to be part of the testing. The research nurses were recruited so that they were two for each geographic region. And they underwent a rigorous five-day training altogether in Santa Monica. I think that was in October. And then they went out to their various markets and geographic regions, and met with the facility and agency staff assessors, and also the directors and such to get in touch and begin to launch the data collection.

The facility and agency staff assessors also went through training. We had actually 14 separate trainings conducted throughout the Fall, so each sort of geographic region or market came to a central location and spent a day...I think it was a day or a day and a half of training together.

So, there were three major types of assessments that we included. We had the communicative admission assessment, the communicative discharge assessment, and a non-communicative assessment protocol. And the non-communicative protocol, this is a little bit of a... we imposed this distinction because we had three data elements that were developed expressly for assessment of patients and residents who are unable to communicate. And we wanted to make sure that we had enough of those.

We also...the main communicative admission protocol has a lot of interview items, and so we separated them out to make sure that we could test all three...the communicative admission and discharge as well as the non-communicative.

A subset of the patients and residents were assessed by assessor pairs. This was for both the communicative and the non-

communicative. And the pairs were the facility and agency staff plus their assigned research nurse. We also had a subset of patients and residents who were assessed repeatedly on admission days three, five, and seven. And this is to evaluate the effect of varying look back periods. So, for the three-five-seven test this included...it just included a subset of the data elements. So, specifically the whole protocol admission assessment was conducted on day three. And then, on days five and seven, the same assessor went back to the same patient and repeated the assessment with a handful of the data elements. And as we go through, I'll indicate which ones those were.

We also had a subset...in addition to this repeat assessment, the protocol was set up so that for the chart review items we could evaluate the effect of looking at the chart based on just admission day one versus three, versus five, versus seven. And so, these were built in to just allow us to evaluate the effect of assessing on different days and how that impacted the performance of the data elements.

Also, near the end of the survey, all facility and agency and research nurse assessors were asked to complete a survey. And the survey asked about their experiences conducting assessment, their sense of the relative assessment burden of the data elements, and their opinion on the clinical utility and relevance of the data elements for their populations.

We also conducted a series of facility/agency staff assessor focus groups near the end of the period in select markets and held a research nurse teleconference. And this actually...Stacy mentioned this earlier, this is really just to assess what's going on on the ground. How do you all feel about asking these questions? What are you hearing from the patients? Did you find that this is useful? Did you come up with...did you see any issues with any of these things?

So, there was a previous special open-door forum where I went over a lot of the...well, actually, Jaime Madrigano went over a lot of the information that we learned from the survey. And in this, we're going to sort of combine the information we got from the survey with what we heard in these focus groups and teleconferences.

Okay, next slide. So, this map lists the 14 markets which were included in the beta test. The markets fell roughly into these three

regions. In the west we had Los Angeles, San Diego, Phoenix, Dallas, and Houston. Centrally we had Kansas City, St. Louis, Nashville, and Chicago. And in the east, we had Boston, Philadelphia, Harrisburg, Durham and Fort Lauderdale.

The markets were selected randomly from a pool of approximately 65 eligible markets. So, the eligibility was determined early on by our statisticians. It was based on several factors, but the one that was most salient, for me anyway, was that the markets had to have enough providers in them to make it worthwhile that we could actually recruit enough participating facilities and agencies in that area. And the goal was to have approximately 15 facilities and agencies per market. We worked hard to get that balance. Also, we were trying to balance that there would be at least one of each type of setting in each market.

So, despite all of our efforts, of course there was variability in the distribution of completed assessments across the 14 markets. This map shows that. So, all the markets made substantial contributions to the field test. But you can see that there was a lot of variability. The completions ranged from a low of 106 in Chicago, and I think 109 in Nashville, but then we got 409 in St. Louis, and over 300 in Fort Lauderdale and Durham and 300 in Boston. And so, it really sort of depended a lot...there were a lot of dependencies actually. But everybody worked really, really hard and we made sure that every market made a substantial contribution to the test.

In terms of our sample sizes, we ended up with a total of 103 participating providers. We actually had a lot more than 143 sign up and our recruitment period was a little extended. And a lot of what happened was that people signed up in July and then by the time they were supposed to go to training in October, we had a new...there was all new management, and everything had changed. And so, a lot of those that signed up with really good intentions, found themselves unable to participate by the time the test came around.

It was a complicated line to walk because if we waited until the last minute to recruit everyone, we wouldn't have had time to recruit enough. And so, we did a lot of work to just keep...carry the recruited facilities along. And in the end, we felt actually really good about the numbers that we got. We had 60 skilled-nursing facilities, 25 long-term care hospitals, 23 in-patient rehab facilities, and 35 home health agencies. And I don't have this here, but the distribution across the markets was nicely disbursed.

And then this bottom slide...the bottom table shows the number of completed assessments. This doesn't have the non-communicative. This is all the communicative assessments. We had over 3,000 admission assessments total which is really remarkable. And then, nearly 1,000 of those were conducted as paired assessments. And then almost 600 patients and residents agreed that we could come back two more times and ask them questions again. And we got a little over 800 discharge assessments. And I think the discharge assessments...we were hoping to get more of those and we really pushed at the end. We found that it didn't quite fit in well with the workflow. The assessors weren't notified with enough lead time to get things together to plan the assessments.

So, that was a little bit of a learning curve with that. And they're also so focused on getting the admissions...it was a lot to manage, but we focused in on it and managed to get enough that we were able to look at the setting level at change over time during the length of stay.

Okay, here are some provider characteristics. We had approximately half were nonprofit and about half were for-profit. We had in terms of the freestanding, it only applies to the IRFs and SNFs and about half of the IRF**s** were freestanding, but the majority of SNFs were freestanding.

And as you can see, although we really tried to get non-metropolitan settings, it was challenging, especially because of the way that we defined the markets. But we did have...although we have 90% metropolitan, we do have a little bit of rural and some representation in micropolitan in and small town. And if anyone's interested in our definitions of micropolitan and small town, I have that documentation.

The other was that this number of beds...the range of the number of beds...and this is another one where there was some requirement that the facilities and agencies that participated had to be of a sufficient size that they could contribute a substantial number of admissions. And so, there weren't a lot of really small sites. But you can see they all range...it looks like maybe 30 was the minimum. It looks like 30 might have been the minimum.

And then the IRFs went up to 881. It was quite a range. So, some of those were really big. I think one of the ones in St. Louis was one of those really big IRFs and that's why we got so many from

there.

And then the nurse-to-bed ratio is another statistic that we document here.

Okay.

Female Audience: **[Inaudible] [00:29:03] (asking whether im taking questions ongoing)**

Maria Edelen: I think...let's get through this first set and then we'll see how we are with time and everything. I kind of like the involvement, but I don't want to get too off track. So, let's see how things go and possibly in the later sections we'll be able to take questions on going.

Okay. So, this slide is just a little bit about the patients and residents that participated. The majority were female. It was a slight majority but higher majority in home health and about half in the LTCHs**.** And of course, as you would expect, they tended to be 65 years or older, but we did have some 10, 11, almost 12% under 65. And a sizable portion were over 90 years old.

The length of stay ranged from about two weeks to about a month on average. And the majority of patients and residents were discharged to home, which is good news. But that majority is really notable among patients in the home health setting. You can see for those who were participating from an LTCH, the majority of those patients tended to go to SNFs afterwards and also to home health.

Okay. So, now we're on slide 21. Now that we've gone through a little bit of the design and the sample description, we're ready to actually get into some results. In this first section, I'm going to just review some of the overall findings and key takeaways that pertain across all the data elements. And then I'll start getting into some of the data element-specific results after that.

Okay. So, in terms of key takeaways, the data elements really performed quite well. And quantitatively, there were really no fatal errors or nothing that really tanked by the data collection. I mean quantitatively, the data elements tended to do really well. They all showed very strong reliability and those results were fairly consistent across settings and across the data elements. There were very few areas of concern. Some of the data elements

didn't do quite as well as others, but there wasn't any really problematic performance and there were no red flags.
In terms of feasibility, there was very little missing data. I'm going to skip to the next slide for a second and then skip back just to show you the missing data. This slide has a lot of information on it. I'm on slide 23. So, I like it because it very clearly conveys the low rates of missing data across all the modules of the assessment protocol. So, it's broken up by...we collected the data on tablets and the tablets you essentially enter one module at a time. And the modules were broken up as listed here.

So, Module A had hearing and vision for example, and of the 3,121 admission assessments, 98% of those assessments had at least one response in the hearing and vision module. So, this is just saying did you go into the module and did you answer at least one question? And you can see that almost all of them...I mean all of the modules are entered over 90% of the time. And then, this All Modules line shows you that nearly 90% have at least one response in every single module.

So, we were actually...it anecdotally sounded like there was a lot of missing data. And then when we saw it, it's like, "No. This is actually really good." So, we were actually very, very pleased that people were diligent. It was tricky too because there were some data elements collected by interviews, others you had to go into the chart, others you might ask the family, or you might have this sort of experience of what's going on in the room with the patient or resident. And so, it was sort of an ongoing activity. And it would have been easy to say, "Oh, I didn't get to that chart review" and just submit the assessment. But the assessors were really diligent and got a lot of this work done.

Okay. The other thing that we don't show here, but then once inside a module there was also missing data. Like you can go into the Hearing and Vision Module and then skip an answer, or maybe the patient or resident didn't want to answer one of the interview questions. So, we have some of that missing data as well, but that was also very, very low.

Okay. I'm going to go back to the previous slide. So, the other big key takeaway has to do with the repeat assessment results, also known as look backs. So, the repeat assessment of the patient interview items on admission days three, five, and seven showed very little variation in responses across days.

So, that means, for example, one of the data elements that was assessed on days three, five and seven was the BIMS (brief interview for mental status) and that includes orientation to time and place. Like what's the date today? And what time is it? If you went in on day three and asked that, and then you went in again on day five and asked that, the patient or resident was pretty able to perform at the same level on both of those days and also on day seven. We didn't see a lot of change over time depending on which day you asked the question.

So, the other ones were like the pain interview and some of the other cognitive assessments. There was also ability to see and hear were assessed repeatedly. The other one, as I mentioned earlier, the recording of the chart information on days one, three, five and seven and we found the majority of the information, if it was going to be in the chart by day seven, it was almost always there on day one. And so, that meant that you could walk in and check on the chart in any time during that timeframe and you're going to get the same answer.

So, this was just...there were a lot of issues around the look back. And I know there's a lot of variability in the requirements across the different settings and so, this was just like okay, if we need to be able to accommodate some of these variabilities, is that going to hurt us? And so, we put this part in the test to try and get a little bit of a handle on what's that going to do and how much flexibility do we have? And so, it's been very encouraging to feel like we do have some flexibility because these results show that, at least in terms of the performance of the data elements, we don't have to really worry about exactly which day the questions are asked.

Okay. So, now I'm going to skip ahead to slide, 24, which I have to smile because of the way that this chart was created. In the spirit of attempting to just give some key takeaways from the beta test at the level of the data elements, my team...I got a small group of project team members from RAND to help me populate this chart. So, the four evaluation criteria at the top were the ones that Stacy mentioned earlier...how did the data element perform in terms of its potential for improving quality? How was its validity and reliability? Is it feasible for use across the post-acute care settings? And does it have potential utility for describing case mix.

And basically, what I asked my team to do was just to rate low, medium, high for each of the data elements. And you know, there's one thing with the empirical information, but as I said, the

empirical information didn't...I mean it was all good news, but it didn't really help us rank order or rule out or highlight per se any of the data elements. And so, I asked the group to sort of think about really everything that they know and everything that they've heard about the data elements across this timeframe and give their best estimate of the performance or the evaluation of the data elements.

And then, there was some variability...we actually rank ordered pretty closely, but there was some variability. And then we got on the phone and it was like a pseudo-Delphi process. We got on the phone and talked about, like why did you give that a three? I gave it a one. I didn't even want to mention numbers though. I don't want this to be taken as...it's not quantitative, but we really did feel like we wanted to get some sense of how the data elements performed relative to one another across these evaluation criteria and this is what we came up with.

Okay. So, now are going to get into the specific results for each of the data elements and the results are going to be presented according to the clinical categories. But first I want to just tell you a little bit about what are the components of the results that I'm going to be sharing. So, for those of you on the phone, I'm now on slide 26. So, for feasibility we're focusing on time to complete. I already mentioned, we also considered the missingness as an aspect of feasibility, but as I already mentioned, all the data almost did really, really well in that category. And so, we're not going to repeat it for every single data element. It was across the board really, really strong.

For inter-rater reliability, we are reporting kappas and percent agreement as a companion statistic. In some cases, for some of the data elements that are low in prevalence like behaviors that are threatening the physical space of other people, some of these are really low prevalence. And for those, the kappa statistic turns out to not be a stable estimate of reliability. And so, in those cases we don't have the kappa statistic. We report only the percent agreement.

And then, as I mentioned earlier, we have three sources of assessor feedback. We have the assessor survey, and we also have the focus groups with the facility and agency staff assessors, as well as the research nurse teleconferences. Then for all of the feasibility and interrater reliability, we did calculate that for each setting separately and overall across all of the settings. And for the most

part, we didn't find a lot of standout differences across settings in any of these results. So, I'm usually going to be giving you just the overall results. And I just want to clarify here that anytime I say overall, it just means that it's combined across all of the four settings.

Okay. And now, we'll get into the results. We're going to start with 'Cognitive Status'...the clinical category. And first, let me orient you to this table because we're going to see a table like this for each of the clinical categories. And basically, what it is is just listing the tested data elements that are included in this category, and then sort of the project history of the data element, like when...how was it involved in our activities? And then how is it included in beta? And finally, if relevant, is the data currently being used in any of the four settings' assessments?

So, for 'Cognitive Status', we have the BIMS (brief interview for mental status) which was included in our first public comment, and also in the proposed rule cycle in fiscal year 2018, calendar year 2018. The BIMS was one of the data elements that was included in the repeat assessment, the three, five, seven. And it's currently assessed in the IRF-PAI for inpatient rehab facilities and MDS for SNFs.

We also have the 'Signs and Symptoms of Delirium', also known as the CAM, which also was in Public Comment One and in the proposed rule cycle, and also was included in the three, five, seven repeat assessment. The CAM is currently assessed in the long-term care data set. This LCDS is one of those acronyms...or not acronyms actually, it's an abbreviation that we have been using on our project, that I don't think is very familiar to you, but that stands for the LTCH CARE data set and also the MDS.

And then we also tested a series of questions about 'Behavioral Signs and Symptoms'. Some of these were presented in Public Comment One and a subset were also included in the fiscal year 2018, calendar year 2018 proposed rule. And then a larger group of behavioral signs and symptoms were assessed in our second Alpha test and also were presented for comment in our Second Public Comment period. And these questions were also included in our three, five, seven lookback test and are currently assessed...a version of them is currently assessed in the MDS.

We had two versions of item sets to assess expression and understanding. I can't remember if both versions were in Public

Comment One or just one of the versions was in Public Comment One, but they were commented on initially. But then we tested these two versions. They were both tested in the three, five, seven.

One of the versions is used just in MDS. That's the three-item version. It asks about speech clarity, and then the ability to understand others, and also the ability to make self understood. And then there's another version that's just two items which is the expression of ideas and wants, and the understanding of verbal content. And so, the expression items in this second version sort of combines ability to speak clearly with the ability to express. And so, the MDS separates it out and the other version keeps it intact.

We wanted to test both of them to see if there were any differences in performance. There were a handful of data elements for which two versions were tested. And in those cases, we randomly split the markets in half and half of the markets got one version and the other half of the markets got the other version. And so, we didn't...we randomized it at the market level just to simplify the training, so that when we trained a market, they were all being trained on the same version.

Okay. So, then the last data element in the 'Cognitive Status' category is the staff assessment of mental status. And this is one of those ones that's developed specifically for patients who are unable to communicate. It was tested in Alpha 2. It was...we solicited comment for it in our Second Public Comment period and it's currently used in the MDS and in IRF-PAI.

So, in this section, I'm going to go through each of these data elements in turn except for the 'Staff Assessment of Mental Status'. That's going to be covered at the very end of the presentation.

Okay. Now I'm on slide 29. We're going to start with the results for the BIMS. So, slide 30 shows just a little bit of the results for the BIMS, but you can see in the top right that it's in the admission protocol on pages six, seven and eight. So, as you can see, it assesses short-term recall with the repetition of three words. It asks for the year month and day. The three words it asks right off the bat "Can you repeat them?" And then at the end it says, "Now what were those three words...do you remember what they were?"

And so, I also should say that this protocol is the beta protocol. It is not how this is going to be standardized. It's not how it is in the

MDS. It's the way that we tested it on our tablets. So, some of the things that are particular to our test are not in, say the MDS version or may or may not be in whatever standardized version ever came out of this. But we had certain constraints by using the tablet and we also had certain design features that we wanted to implement. And so, this is really specific to our testing.

In any event, the BIMS took approximately two and a half minutes...2.2 minutes to complete on average. We didn't have a lot of variability in the time to complete across the settings and it showed excellent reliability. So, in the paired assessments, the raters tended to agree over 90% of the time...almost 95-100% of the time. And the overall kappa range was also quite high.

So, the kappa, just to orient you, you really want to see kappas around .7 or .8. That conveys pretty high agreement. Point 4 to .6 or .7 is reasonable. And lower than .4 is really not what we're looking for.

In terms of our assessor feedback...let me also orient you to this slide. We're going to see a lot of slides like this. I have a version of this for every data element. And we have little icons for the three sources of information. And the icon indicating the source...so for example, for the clinical utility, we heard about that from the assessor survey. So, that's the one that's bright. We also have an icon for the facility and agency staff focus groups and for the research nurse teleconference. We got a lot of information from a lot of different sources and this is really just kind of a synthesis of all of that.

So, the BIMS...we got a lot of support for the BIMS. It was rated very high in clinical utility on the survey. The research nurses found it to be really helpful to assess cognition consistently across post acute care and over time. They just found it to be a really clinically useful assessment to conduct. Also, in the survey, it was rated as having pretty low burden...relatively low burden, especially for home health. And then the staff assessors and the focus groups mentioned that a lot of staff are already really familiar with the BIMS or with assessments like the BIMS and so, they saw it as not a heavy lift to orient people to assessing cognitive status in this way. So, that was a strong vote of support from them.

There were a few challenges about the BIMS. One that came up was that if you used the BIMS over and over again on the same

patient, they start remembering bed, sock and blue. And I mean, I know bed, sock and blue. I think a lot of us know bed, sock, and blue. So, it's a little bit of a problem. Part of it is because we have that three, five, seven, so especially those patients. They were saying, "Well, you just asked me that two days ago." It's not clear that it's a huge problem because we've talked about it a fair amount, actually. And even for those who acknowledge that some patients might remember, they say, "Well, so we're testing their long-term memory instead of their short-term memory. It still says something that they can remember it even if it was from last week."

So, it's possible that somebody might want to do some research and come up with some other words, but it's not a simple thing because you want them to be monosyllabic, and you want them to be able to be clearly understood. And in fact, this ability to enunciate the words also up came up as an issue. So, for people who didn't speak clearly, or didn't speak loudly, or perhaps had English as a second language the assessors, if they weren't able to say the words with the appropriate enunciation, then the residents and patients couldn't play it back because they didn't hear it well. So, there's some burden on the staff assessors to make sure that they enunciate very, very clearly in order for this to work out.

But all in all, those seem to be challenges that can either be overcome or need not be worried about too, too much.

The Confusion Assessment Method or CAM is next one. And that's on pages seven and eight. So, it's meant to directly follow the BIMS. It's supposed to be assessing the signs and symptoms of delirium. It evaluates whether there's an acute onset of mental status change, whether there was inattention, difficulty focusing, disorganized thinking, and altered level of consciousness. And so, this one took just under a minute and a half up to complete on average to do the whole item set. And the percent agreement was excellent for this one.

The next slide...now we're on slide 34. The CAM was a funny one in terms of the assessor feedback. It kind of went under the radar. It wasn't really problematic. We didn't hear anything negative about the CAM. We didn't hear about challenges. Nobody said, "Oh, that was so hard to figure out. It was hard to complete." Nobody said anything like that, but they also didn't say a lot about it in support either. So, we were a little at a loss, but we did go to the assessor survey and saw that it was ranked moderately high in

terms of its clinical utility. So, the survey asked the assessors to rate each of the data elements in turn. So, we made sure we got at least some information for all of them. The CAM was seen as moderately high in clinical utility and relatively low in burden. So, we did get sufficient support from the assessment...the assessor survey.

Okay. 'Behavioral Signs and Symptoms' is the next cognitive status. Oh, I'm running behind. The 'Behavioral Signs and Symptoms' are on pages 47-49. I think this is the last one. Right. So, this assesses whether the patient or resident exhibited any physical symptoms or verbal, or other behavioral symptoms directed at others or directed at themselves. And then, if there were any of those symptoms, was there any impact? Was there an impact on the patient or resident? Was there an impact on others? And finally, it assesses whether the patient or resident had behaviors that caused them to reject their evaluation or reject their care.

This one was a really low prevalence question. It took 1.4 minutes overall to complete. The percent agreement was very, very high. We didn't get kappas for this one because it was such a low prevalence data element. And in terms of assessor feedback, it was seen in both the focus groups and the teleconferences as having really high clinical utility. They thought that it was very important for effective transfers across post acute care settings. And similar to the MDS, the focus groups were mentioning that staff typically tracks this type of behavior and that it wouldn't be that hard to do that in the standardized fashion. So, they saw this as not a heavy left to get this one in place.

There were some challenges, especially in home health. They found that they just didn't know enough about the patient on the first visit to be able to complete this. And then there was some discussion about certain staff possibly being reluctant to document this information because they are concerned that it would hamper their ability to transfer the patient to another setting. So, we weren't sure how seriously to take that, but it was something that people mentioned.

Okay. Now I have 'Expression and Understanding' which is on pages four and five. And this is the three-item set and the two-item set that I mentioned earlier. The three-item set took just under a minute to complete overall and the two-item set very close with 0.7 minutes overall to complete. The percent agreement was excellent

for both versions. The three-item set was in the 90s and the two-item set was in the high 80% agreement. And the kappas if they were calculated, were moderate for the three-item set and a little lower for the two-item set.

In terms of assessor feedback, we didn't get any challenges or concerns for these items set. Everybody saw it as having really high clinical utility especially in the LTCHs and SNFs. People saw it as important for facilitating patient transfer. They found that it assisted with interpersonal connection with the patient resident and also reported it as having relatively low burden, especially for home health and inpatient rehab.

Okay. So, now we're ready for questions. How are we doing that, Emily?

Emily Chen: **[Inaudible] [00:59:15]**

Maria Edelen: Operator?

Male Operator: Yes. And again, ladies and gentlemen, today's conference is being recorded. If you have any objections, you may disconnect. The presentation is not intended for the press and remarks are not considered on the record. If you are a member of the press you my listen in, but please refrain from asking questions. If you have any inquiries, you can contact CMS at press@cms.hhs.gov. And for the rest of the audience, if you'd like to register a question, please press the 1 followed by the 4 on your telephone and you'll hear a three-tone prompt to acknowledge your request. And give us just a minute for the first question on the phones.

Emily Chen: Okay. And I think we'll alternate between someone in the room and then we'll go to someone on the phone. So, in the room...

Carol Carter: Hi, this is Carol Carter at MedPack. I had two questions. The first has to do with, when you were looking at the utility for describing case mix, I wondered when you're thinking about potential use for payment models, whether you were thinking about the potential for upcoding because the commission is increasingly concerned about the patient assessment items being upcoded from what we're seeing in a couple of the settings. That's my first question.

Stacy Mandl: It feels like a CMS question. So, we share the concern. And we are looking...obviously, we have talked about it before in previous rule making, but definitely looking at the feasibility of data

validation program that looks at data accuracy because of the importance of having reliable data that's accurate and actually clearly and completely, consistently reflecting individuals is very important for everybody. Not just for CMS purposes and not just for payment, but for real patient care.

[Unidentified Person]: And to add to what Stacy said, in our historical studies of accuracy with regard to payment items specifically in skilled-nursing facilities, it was under a contract called the Data Accuracy and Verification...I don't remember. Anyway, it sort of balanced itself out. There was as much under coding as there was upcoding. That's old data of course, and that was limited to just whether that trend existed anywhere else, I don't know.

Stacy Mandl: Yeah. That doesn't surprise me given there's so much emphasis on doing therapy that I don't think there's so much emphasis on upcoding functions. So, I don't think that not seeing it in the SNF is particularly representative of the other PAC settings.

Carol Carter: My other question had to do with the sample because it's not particularly representative of what you see in the industry either by provider mix or patient mix. And I'm wondering whether that was purposeful as kind of in over sample for IRFs and LTCHs and whether that was done on purpose or just sort of what you got. And probably, more importantly, what difference do you think it would make?

Maria Edelen: Sure. So, we definitely over-sampled the IRF. I mean the LTCHs you have to if you want to get them in there. If we wanted to not over-sample the LTCH you would have to have many, many more SNFs and home health just to balance it out. So, what we did was we determined the minimum number of facilities that we needed for our power. And that's how we decided...that's how we determined the number of LTCHs. They were our minimum.

And then the other ones were more...and the IRFs were certainly difficult to recruit. They're not as difficult as the LTCHs but also difficult. And then for the home health and the SNF it was more capacity. We wanted to have more because there are obviously many, many more in the country proportionally. So, we wanted to have more of them to sort of reflect that, but to have it at the right multiple would have been just an enormous task and not worthwhile.

We are considering doing some weighting. Depending on the

types of analyses, but for things like reliability the main thing is to have enough to get stable estimates. If we end up doing analyses where we're trying to pinpoint rates, or trends, or say something about the national sample or the national population at large for any reason, I don't know that we plan to do that necessarily, but we certainly have the wherewithal and the resources to do the weighting to stretch it out to be nationally representative.

The other thing that we're going to do in our beta report is evaluate the lack of comparability. So, I showed you sort of the urban and rural mix and the free-standing profit mix, etc. We're looking at the national data and comparing those rates to see if there are large differences in what we have versus what's in the national data.

We're also, even within our sample, looking to see whether the subset that received the interrater reliability paired assessments is reflective of the overall admission assessment sample. So, we're doing a lot of those checks. That's one reason why we're calling these results preliminary. It's all the data, but it's not all the results. And the beta report will go into all of this and provide sort of rationale and logistics and comparisons for the representative aspect of things.

The other thing to keep in mind is that you have to balance feasibility with representativeness. And so, especially the urban rural mixes, which is unfortunate, but we had to make sure that this was feasible to have research nurses who could access the locations. And we did our best to get a few rural places in.

And also, the performance of the data elements themselves...I personally feel like that type of statistic generalizes pretty well. I think if you're talking about rates of some clinical condition or something about quality, or the actual sort of content of the data element, and you're trying to make a statement about what's happening in the country, that's when it's really, really more important or critically important to have the representativeness. And if we are going to make statements like that, we'll certainly do our best to weight them or caveat them so that they don't get misrepresented or misinterpreted.

Emily Chen: **From the phone?**

Maria Edelen: Sure. Do we have anything from the phone?

Male Operator: We do. We have a question from Tammy Haller. Please go ahead.

Tammy Haller: Hi. I'm wondering for those patients who were discharged to hospice, are there any assessment data or indicators that might be extrapolated that could be helpful to hospices to determine possible eligibility or life expectancy?

Maria Edelen: So, I think I heard you say hospice and we didn't have...hospice is not one of the settings in our beta test. We had home health.

Tammy Haller: Yes. I understood that, but you listed in your list of discharges, there was a certain percentage discharged to hospice. And I was just wondering if any of the discharge data might, for those patients who were discharged to hospice, might be extrapolated to be used as a potential indicator for appropriateness for admission to hospice. Does that make more sense?

Maria Edelen: Yes. That makes a lot of sense. How many... well...this is a great idea actually. I'm looking at Anthony who does our analyses. It's a small percentage, but I really, really like the idea of seeing whether any of the data elements at admission or discharge might be related to being discharged to hospice. So, we will definitely do that, and it'll be in the beta report.

Tammy Haller: Thank you. I appreciate that.

Maria Edelen: Thank you!

Tammy Haller: I'm glad you see the value that it could bring.

Maria Edelen: Yes. Oh, another question from the room.

Teresa Lee: Teresa Lee with the BNA Health Group. I had a question about a comment you made on the...let me see, what slide number is this? It's the slide on completed assessments for each module. And you made mention of the fact that the percentages are quite high and that it seemed inconsistent with anecdotal information that you picked up. And I guess my question is what kind of anecdotal information? Is that from focus groups? Or was that just sort of from talking with people participating?

Maria Edelen: That's a good question. It was more...every week we got on the phone with our survey coordinator, and with the data collectors, and with some research nurses, and with the project manager, and the people who were talking with people on the ground. We were sort of trying to get a lay of the land. What's going on out there? And it was not so much Chicken Little, but almost. We were

really...we were in the business of trying to determine what the issues were and resolve them. And so, when we were on the phone it sounded like...for example, the survey coordinator would say, "Well, I'm getting a lot of submissions that are incomplete..." I wasn't looking on a daily basis to see how incomplete it was. And so, based on these conversations, I had some concerns that it was going to be a lot worse. That's all that I meant. And it was purely because...it was really in an effort to troubleshoot. That we were focusing on our shortcomings. And then ultimately, I guess that focus made it not so short which is good news.

Oh, another one from the phone.

Male: Yes, we have on the line Deb Head. Please go ahead

Deb Head: Hi. This is Deb Head and I do have some concerns about the feasibility or the sampling sizes. Were patients chosen based on characteristics, or were you seeing a full scope of the patients that the facility would have?

Maria Edelen: That's a really good question and a detail that I didn't convey. So, the intent was to include or enroll every admitted patient or resident that was admitted during the field period. I think some of them were missed, but if they were missed, it was due to logistics, not having to do with any patient characteristics. There were some who were not eligible if it had to do with safety, like if they were in quarantine, but for the most part everybody was eligible if they were admitted and the intent was to enroll everybody who was admitted during the field period, not select them.

Deb Head: That's good to hear that it was all-inclusive because otherwise you don't want cherry picking, maybe patients that were easier or less difficult to actually perform the assessment on and not getting that full sampling. Thank you.

Maria Edelen: Right. Sure. Anything else from the room?

Cynthia Morton: Cynthia Morton with National Association for the Support of Long-Term Care. Are the data elements that you tested for cognitive status meant to reflect cognitive status of the patient? Or were you only testing certain parts of cognitive status?

Maria Edelen: That's a good question. Cognitive status is a really diverse, multi-dimensional concept and we looked into a lot of different assessments for cognitive status. We had some performance

assessments that we tested, we did a clock draw, we did the menu test, we were considering medication management...can they manage their pills? There were a lot of different things that we tested. The functional cognition was one of those things that we...those were basically what I was just describing, that we really wanted to try and get in there to sort of broaden the scope. There's also some patient safety...cognition around safety that we thought might be important to evaluate.

And I think this is work that's still needs to happen. For the time being, it's pretty basic short-term memory, orientation to place and time, and assessing delirium and ability to express. So, it's very basic cognition. We do need more work on that. Do we have another one from the phone?

Male Operator: No other questions.

Maria Edelen: Okay. Can we take one in the back here?

Andrew Barrett: Hi there. Andrew Barrett from Encompass Health. Thanks for presenting this information. Did you all do any sort of meta-data analysis between the types of certifications or trainings that the people who completed the assessments had comparing those two results? The simplest example I guess is with time thinking that it might take people who have less experience or come from a different training background to take longer for the assessments. But you can imagine applying that principle really a across all the different items. Was there any sort of meta-analysis done about who was actually carrying out the assessments?

Maria Edelen: So, yeah. Good question. The time estimates...wait, let me back up. The assessors were primarily registered nurses at the facilities and agencies and they were nominated by the providers. They didn't have any specific training per se. Some of them might have been MDS coordinators. In our assessor survey, I think we asked a little bit about them, but we wanted that to be somewhat anonymous, so we didn't want to really get into a lot of personal characteristics about the respondent.

But we have a sense that they tended to be registered nurses without a lot of experience. They all had the day of training...the in-person training that we conducted. But their experience with the assessment, of course, makes them go quicker. So the assessments at the beginning of the field period maybe took a little bit longer, and then towards the end of the field period it

maybe went more quickly. I think...Anthony, do you know if we've done that sort of looking at the quartiles of time, etc?

Anthony Rodriguez: We did. So... oh, I'm sorry. So, we did look into time patterns and distributions to kind of account for a practice effect component where let's say in the beginning it took longer periods of time. And so, we applied a standard method of looking at the quartiles to be able to trim off extremes so that we would have the best and most precise estimate of what the time is to complete.

Maria Edelen: Should we do... We're a little over, but I thought we might want to do one more question. Oh, there are two in the back. Well, let's do one more in the back and then we have another couple of...I think we're going to end up with more time in the second section, so we'll get to you later.

Tray Hillman: Tray Hillman from UDS Mar. Just a couple quick questions. You mentioned a beta report and I'm sure everyone here is probably wondering what is the timeframe for that to be delivered? And perhaps you're going to mention that later on in the content, so if you want to choose to skip that, that's fine.

Maria Edelen: Yes. We are drafting that, but before we make that publicly available, we have to dot all the 'i's and cross all the 't's and be really, really clear on exactly what's ready to go out there. So, I think the time frame for that is still sort of under discussion.

Tray Hillman: And then the other part of this was, you're mentioning the validity of a part of the reliability and validity as a potential thing that you are testing –

Maria Edelen: Yes.

Tray Hillman: – or evaluating and to date it's been qualitative in nature if I'm understanding that slide that you had previously? As well as utilization for case mix. Will predicted validity...construct validity, those types of validity tests be done on the data, especially since in the most recent IRF PPS rule the BIMS, for example, was considered for payment, but their validity test showed that it didn't complement the payment system. So, in evaluating, will you be doing some kind of predictive validity test to see whether each of these data elements do have the capacity to define case mix, or to explain cost, or length of stay or some of those other outcomes that are available in the post-acute care setting.

Maria Edelen: That's a really good question. We have...I'm going to start. You might want to add to this, but I'll start by saying that we have...we're getting assessment data for those patients and residents who are in our beta test. And so, we are able to...we're just starting to look at, for example, the validity of the BIMS in terms of...hopefully it's related to stroke. Right? We want people who have suffered from stroke to have a lower BIMS score than people who have not. So, really simple known groups comparisons type validity. Like is this measuring what it's supposed to be measuring?

So, we have the assessment data for that. To get into really more clinically relevant validity, we could try to get the claims data. That's going to take...that's sort of another chapter of data request and merging and not something that we are focusing on now, but we actually talk about quite a bit doing. And I think Stacy is ready to talk so I'm going to leave it at that.

Stacy Mandl: That's an interesting question. It's definitely very reimbursement sort of focused, I think. And I think it's important to really understand that the body of this work pertains to much more than just payment, right? As a clinician knowing somebody's cognitive status quickly is very important for care planning, clinical decision support, information exchange of how somebody sort of looks at the time of discharge, when you're receiving them in a timely manner.

So, I hear your question and I appreciate it. I just also kind of want to draw attention to the fact that standardized assessment data is also for direct patient care as well, hopefully. I mean that's part of what the intent was in looking at sort of those principles. Information that clinicians need to know at the time that they are caring for them and when they are transitioning them over to someone else. So, thanks.

Maria Edelen: So, how do we want to do this agenda-wise. Do you want to take a... should we keep it as ten minutes still? Or do you want to just do five?

Emily Chen: Let's do five.

Maria Edelen: Okay. So, we'll take a quick break then, and will start back up at 1:35 PM. Okay. [**Recording off]**

Maria Edelen: Okay. I think we're going to start back up. Okay. Great. So, the next category of data elements that I'm going to describe are those that assess 'Mental Status'. They include the PHQ-2-9 which is an assessment of depressed mood. And the PHQ-2-9, aspects of that were included in Public Comment One and in the proposed rule. And I should say that we noticed that we got the years wrong for the...it should be fiscal year 2017, calendar year 2018 proposed rule. So, throughout this whole presentation that needs to be corrected. Sorry about that.

Male Voice: **Correcting Maria**

Maria Edelen: Eighteen and '18? It's '18 and '18. Okay. Because it used to be just the fiscal year and then we added the calendar year and we added it wrong. Yes. Yes. Okay. and the PHQ-2-9 was also tested in Alpha 1. The PHQ-2 is currently assessed in the OASIS and the PHQ-9 is included in the MDS.

We also evaluated two PROMIS data elements. One was PROMIS Depression and the other was PROMIS Anxiety. And PROMIS stands for Patient Reported Outcomes Measurement Information System. It's a large NIH initiative to evaluate patient-reported outcomes in a more standardized fashion. It's a huge measurement or assessment initiative that we wanted to try to incorporate into our work. And so, among the domains that are assessed by PROMIS, these were two that were good candidates for post-acute care assessment. For the PROMIS Depression we sought input from the technical expert panel and also solicited input from stakeholders via email.

Two versions of PROMIS Depression and PROMIS Anxiety were tested in beta. One used the standard PROMIS format, which mostly was asking about their experiences in the past seven days. And we tested an alternative version to be more aligned with post-acute care assessment asking about patients' and residents' experiences in the last three days. PROMIS Anxiety was...we incorporated that a little bit earlier on into our work and it was tested in Alpha 2 and was also included in our Second Public Comment.

And finally, we have the 'Staff Assessment of Mood' which was the observational version of the PHQ. And that, I'm not going to talk about right now, but we'll get to it at the end. We tested it in Alpha 2 and it was included in Public Comment 2 and it's also assessed currently in the MDS.

So, the first one we want to look at is the Patient Health Questionnaire the 2 to 9 item. And the reason we call this the 2 to 9, if you look on pages 23 to 27, essentially it is the nine item PHQ. It assesses all the symptoms, or it includes assessment of all the symptoms. What we tried to do with this so-called hybrid version, is that we're using the first two items, which are the cardinal symptoms of depression, to screen out patients and residents who really are not at risk at all according to their answers to the first two questions.

For those who say yes half or more of the days for either the first question or the second question...so, if they are saying more than half of the days they've been bothered by having little interest or pleasure, or more than half the days they've been feeling down, depressed, or hopeless, we consider them somewhat at risk and we continue on and ask them the remaining seven questions. If they say, "No, I've not been bothered with having little interest or I've been bothered a little bit, but less than half the days," we consider them not at risk and they screen out.

So, that actually has a nice...makes a nice dent in the burden in time to complete. As you can see, overall it took 2.3 minutes to complete the PHQ-2-9 for all patients and residents. The gain isn't as large as you would hope because you basically, no matter what, where you would ask them two questions, or you would ask them nine questions, you still have to orient them to the assessment. You still have to engage with them. You have to get into it with them. You have to talk about your mood.

So, even though you're only asking two questions, it still takes a little over a minute and a half to do that. But for those who complete and go all the way on to do all nine, it takes four minutes. Now again, that four-minute...it's not going to take four minutes for everybody who goes through the nine because that four minutes is for people who are at risk and who are relatively more depressed. So, they're saying 'yes' to more things. So, if you're going to do the PHQ-9 on everybody, the expectation is that that time estimate would be less than four.

But in any case, for our beta test, for those who skipped out it took less than two minutes to screen them out. And wasn't it...Anthony, wasn't it about 75% of patients who screened out? Right. So, for the 25% who show some risk according to their answers to the first two questions, you go on and ask them all nine and it took an

average of four minutes.

The reliability for this version of the PHQ was excellent. The percent agreement was very high and the kappas were also extremely high.

On Slide 46 you see the assessor feedback. Both the staff...the clinical assessor... staff assessors and the research nurses saw this as having very high clinical utility and thought it was very important to assess mood. We heard from all of our sources in terms of challenges and concerns that it was a relatively burdensome assessment. It takes some time, it's personal, sometimes the assessors didn't feel always comfortable getting into it.

This was actually a barrier when MDS 3.0 came out. They found that they could overcome that just by sort of getting the nurses more used to having these kinds of discussions. And they ultimately found that it actually was really helpful. So, this burdensome...the burden issue appears to be somewhat overcome with practice in making this more of a routine. And also, really good training.

Some of our assessors had some issues with the wording of some of the items...not wanting to ask about feelings of hopelessness. It's sometimes hard for the patients to understand. So, there was some issue around that. And the other was the two-week look back was a little confusing or maybe too long for some of the patients and residents.

For the PHQ, it's a standardized assessment and it needs to be asked in two weeks because it relates to a depression diagnosis. And for symptoms to matter or for a diagnosis to be indicated, the symptoms have to be present...have to be asked about over the last two weeks. So, that's why we have that two-week look back. So, even though it's somewhat challenging, it's essential for it to be diagnostically relevant. And that's why we keep it in there.

PROMIS depression...if you look on pages 28 to 30 in the protocol, you can see these questions. They are similar, but they are asked about in a slightly different way. People are asked on a Likert scale. With the PHQ-2-9 we ask first if the symptom is present and then, if it is present, how often in the past two weeks. These are not directly related to obtaining a diagnosis or a risk for diagnosis. These are really just...this is sort of like a depression

symptom inventory. And one of the versions, as I said earlier, was based on the past seven days and the other version is based on the past three days. I felt worthless, I felt I had nothing to look forward to, I felt helpless, I felt sad, I felt lonely, I felt depressed...that's six of them. I think there are eight. I felt I had no reason for living. I felt hopeless. That's the eight. It took 2.2 minutes overall to complete the PROMIS Depression questions... the eight questions. The percent agreement was excellent as was the kappa. The overall range was from 0.96 to 0.99. So, in terms of the...statistically it performed quite well.

So, one of the things that we found in terms of support for the depression... PROMIS Depression was that it doesn't require the patient or resident to say, "Yes, I'm depressed." And so, the staff assessors found that to be somewhat valuable...that you didn't have to lock them into saying they were depressed. And it might be a little bit easier for them to work through those items.

Similar to the PHQ-9, it was burdensome I think for the same kinds of reasons. It's hard to get into these conversations I think for a lot of the assessors. Especially if they're not used to doing that.

And then, the wording in the introduction is a little off from what's actually...you're asking about the past seven days, but then the intro... some people found not really to align well with what the questions were actually asking. But again, these are things that could be modified. They're not a huge concern. They are things that could either be modified by changing the data element or also by working on the training a little bit.

Okay. PROMIS Anxiety is on pages 31 to 33. This is also eight questions. A lot of it was about worry, or trouble paying attention, feeling nervous, hyper focus on anxiety, needing help, feeling panicked. So, these eight anxiety symptoms took approximately 2.2 minutes on average to complete. The percent agreement on this, similar to the depression items, is very high. It was between 97 and 99% agreement for the items. And the kappas were also extremely high ranging from 0.96 to 0.99.

Slide 52 shows the assessor feedback on the PROMIS Anxiety questions. Similar to the PROMIS Depression, the facility staff especially thought that it was good that the wording didn't require patients to self-identify as anxious. There was some anecdotal feedback that...just asking the patients and residents about anxiety

made them a little put off. Like, "Wait, should I be anxious?" And wondering why they were being asked these questions and so, I think it was nice that they didn't have to actually self-identify as anxious in order to endorse some of these questions. And they found it, in the survey...the assessor's survey, the anxiety data element was rated as moderately clinically useful.

This is seen as somewhat burdensome in terms of the length. We tested it in Alpha 2. It had 10 items and the length was seen as a pretty significant burden. And so, we went down from 10 to 8 for the beta test to try to mitigate those concerns a little bit.

Next, we have the 'Pain Interview'. Well, we included the pain interview. We also included the staff assessment of pain. I'm going to talk about the staff assessment at the end. The pain interview asked about pain presence, frequency, severity, the effect on sleep, interference with therapy and non-therapy related activities, and also pain relief. Some of the pain items were included in our first public comment. The whole pain interview was tested in Alpha 1 and I think revised a little bit. And also included in Public Comment 2.

We tested two versions of the pain interview in beta. One asked about pain in the last three days to align with the rest of the beta protocol. And the other asked about pain in the last five days to align with what is currently being done in the MDS. This again was...if we were going to move forward with one or more of these pain items, was it going to really matter how we ask the question, or do we have some flexibility in that? So, because it's already being asked in the MDS...a lot of these questions are being asked in the MDS.

So, the pain interview is on pages 17 to 19. And this is another one with a skip pattern. The first thing is determining whether or not the patient has had any pain or hurting at any time in either the past three days or the past five days. If they say "No", that's the end of the pain interview. If they say "Yes", then they go on to answer the remainder of the pain questions.

So, I think this was sort of a reverse. It was like 75% of the patients and residents did have pain. So, only 25% said, "No pain" and didn't complete the rest of the interview. The time obviously, is shorter for patients and residents who didn't have any pain. They were asked one question and then they were done. But overall, it took just over two and a half minutes to complete this

interview. The percent agreement was excellent. The kappas were very high...0.93 to 0.99 for all these questions. And the assessor feedback was very supportive. All of the assessors found this to be of very high clinical utility, particularly the items that were assessing function and ability to sleep. So, their interference with the ability to carry out their therapy and conduct other activities. They also thought of it as being a very low clinical burden. So, we had a lot of support for these items.

Oh, okay. Next, we have 'Impairments'. And for impairments we had first the questions assessing ability to hear and ability to see. We solicited comments on them in Public Comment 1 and also in the proposed rule. They were included in the proposed rule fiscal year 2018. They were being used...the ability to hear was in the OASIS but is not anymore. And both are in the MDS. And the ability to see question is also in the OASIS.

We also have for 'Impairments'...we have incontinence, we were asking about appliance use and frequency of incontinent events. That was more like an observational chart review set of items. We also had some questions directly to the patient about their perceived burden about their incontinence. And those were tested in Alpha 1...and put out for public comment in Public Comment 2.

The chart review piece of it was recorded on admission days one, three, five, and seven and on discharge date as well as discharge date minus 2. Many of the appliance use items are also currently being collected in the OASIS and in the MDS. The frequency of incontinent events is collected across the board in all four with slight variations, but basically the same question.

So, for 'Hearing and Vision', you can see these questions on page three of the protocol. These are really straightforward. They don't involve asking a patient. It's really just based on the assessor's experience of the patient. What is the patient's' or resident's ability to hear from adequate to highly impaired? Is the patient or resident able to see in adequate light...ability to see in adequate light from adequate to severely impaired. This test took just over half a minute overall to complete, so 0.3 minutes per data element.

The ability to hear percent agreement was quite high. The kappa was acceptable but not great. This is a very finely grained Likert scale and where there was disagreement it was usually between...like in the weeds. Like between 2 and 3. Some people said two, and some people said three, but we didn't see a lot where

one person said, “Oh, this person can totally see” and someone else said they couldn’t. It was more in the gradations where we saw some disagreement. And even with this highly fine-tuned scale, the kappas were acceptable.

In terms of the assessor feedback, everybody found this very, very clinically useful and important for facilitating effective transfer and for assessing patients at baseline. And also, important to establish in order to then conduct the rest of the interview. It was seen as having the lowest burden of all the data elements that we tested. And they found that it was easiest to assess in home health.

The continence questions are on pages 41 through 46. I think these are just the...so, if you look on page 41, you first see the few interview questions: have you experienced any incontinent events? And how big of a problem or burden have these events been? So, they do that for both the bladder and the bowel. And then there’s also a series of questions that are asked and completed based on documentation in the chart.

So, the chart review portion took three and a half minutes to complete overall for the entire incontinence data element. The interview took just under a minute and a half. The chart review percent agreement was pretty good. It ranged from 74 to 100%. In terms of the kappas, we couldn’t compute them in all cases because some of the rates...the prevalence rates for certain device use, for example, were pretty low. But where we could compute kappa, they were reasonable...0.55 to 0.79.

The interview agreement was much better...much higher. The overall range was from 98 to 99%. The kappas range from 0.96 to 0.98.

In terms of assessor feedback, the focus group discussion indicated that the staff assessors found it clinically relevant for decision-making and really important for facilitating transfer, especially in a patient’s or resident’s ability to go back home.

There were some challenges in assessing these data elements. All the assessors, both the research nurses and the facility and agency staff assessors, found that they had to consult multiple sources to get the answer. They had to not just look in the chart but also ask the family. And sometimes what the patient or resident said didn’t really match what they were seeing in the chart. They found that the information in the chart was sometimes not adequate to

evaluate. As I said, there was some incongruity between the various sources in terms of whether or not this was a problem or whether it was occurring. And there was...this is one where there really was some variability according to setting. In some settings, it was really clearly documented and in others it was not.

Now we're going to discuss the 'Special Services Treatments and Interventions'. This data element includes a list of nutritional approaches including the existence of an IV or feeding tube and special diets. Those were put up for public comment in Public Comment 1 and some were included in the proposed rule. All of these special services and treatments were recorded on the one, three, five, seven chart reviews. And also, these were included and evaluated on discharge and also looking two days before discharge.

We also had these 'Services and Treatments' including cancer treatment, respiratory treatments and other treatments. There are some of these data elements, in some form, are included in many of these assessments. As you can see, on the right-hand side, there are minor variations in the way that these are asked and in the places that they are asked in the protocols.

And you can see these questions on pages 54 to 56. So, when we look at these, I mean this is a good place to note that this one, three, five, seven really is there for testing purposes only. It makes it look like a pretty burdensome item, but it conceivably could be collected with just one column and one check box.

So, this data element only took 3.3 minutes overall to complete. There were very few differences across settings in terms of the time to complete. The agreement was pretty high...79 to 100% for these data elements. The kappas were a little spotty. They were, again, low frequency events. We couldn't get a kappa for those. But where we did get kappas, there were within the acceptable range.

In terms of assessor feedback, the research nurses in particular noted that they felt it was really important to track all of these special services and treatments. Especially to facilitate transfers. And the assessor survey indicated a lot of support for the clinical utility of these although the IRF...the assessors from the IRFs did not think that it was very clinically useful.

There were some challenges like the other chart review items. It was just a little bit difficult to find the information in the chart.

And that varied across the systems that were being used in the various settings. It also depended I think...in some cases it was a little bit harder for the research nurses because they have to orient themselves to the system that was being used in a given setting, whereas the staff assessors were used to their system because it was their system.

Both the research nurses and the facility and agency staff mentioned that there was very poor documentation about these types of services and treatments in home health. And again, the IRF assessors found these to be of not very strong clinical utility.

Wow, I got ahead. Now I'm ahead of myself. Okay. I think we're going to stop there and take some questions. Do we have any questions from the phone?

Male Operator: No questions, but again, to cue up please press 1,4 now.

Maria Edelen: Okay. Go ahead.

Carol Carter: This is Carol again. I have two questions. One, in terms of the analysis when you get there, I was intrigued by something you said about...I think it was in talking to patients about depression. Whether that would be sensitive to staffing levels, where with higher staffing levels you would think the staff would have more time to spend with patients, so you may be able to elicit more accurate responses from patients. So, just sort of an idea of looking at, not just by setting, I also think we might see some differences by ownership. And so, I would look at that. But also, my mind went to if there are differences by ownership, am I really thinking there's differences by staffing levels or mix of staffing. So, you might want to look at that.

The other reaction I had in some of the clinical relevance for some of the settings like home health and IRF, I wouldn't expect to see very many of the high-end service provisions. So, they may not find it relevant because they actually just don't see those patients. So, it doesn't explain their patients because they're just not there.

And so, I think you need to kind of have your hat on for relevance for like where are these patients in the setting before I would take a lot of...I guess I would need to interpret those results with that in mind of where are you actually seeing patients to know whether it's relevant for the clinical staff or not.

Maria Edelen: So, after the first comment that you made, I agree that nursing staff with more time may be able to evaluate depression maybe more thoroughly or more rigorously, but this is the value of having a standardized assessment. Everybody went in and asked the same questions and took an amount of time and got an answer. And that answer then relates to risk for a depressive diagnosis. And if that risk is high, then there is a next step that's involved, and somebody comes in and says, "This patient or resident needs some attention."

And so, in terms of standardized assessment, it took the time that it takes. Having more time...I'm not sure if I'm being really clear, but the assessment itself takes an amount of time. That's the burden. And so, the question is really does your average SNF staff have the capacity to take on two and a half minutes of questions about depression.

Carol Carter: So, then to me the question is if you don't really have much of a relationship with the patients, then maybe in your two and a half minutes it took you to get your answers, they're actually not reflecting really what the patient is thinking or feeling because they don't really have the relationship. That's what I'm thinking. I understand the time to complete. That's not what I'm talking about. It's whether you're capturing an honest response from the patient because you don't have much of a relationship.

Maria Edelen: So, it goes to rapport. Yeah. But again, it's not a diagnostic interview. It's, "Have you had this and how often have you?" So, sure if you've got somebody who's really closed off and you have no rapport with them, they're just going to say no and you're never going to know. But if you have even a little bit of rapport, and they're able to indicate at least a little bit of trouble, then somebody else is going to come in and be able to build up that rapport and be able to get to the bottom of it.

And the second piece, I agree with you. And of course, this is something...is this relevant across settings? I mean that's one of our big questions, right? And that's why we tested across all four settings. The frequencies of some of the special services and treatments are very low in some settings. But they're not zero. I don't think there was one special service for treatment that was –

Carol Carter: **[inaudible] [01:56:06]**

Maria Edelen: Right. Right.

Stacy Mandl: I just don't like it. I just wanted to add to that because I appreciate that. When you're looking at individuals who have medical complexities and multiple chronic conditions and those that really need the eye as well as their risk adjustment needs, those factors...asking the questions about whether they are on a vent, what their nutritional program is...those are critical pieces of information. And what's nice about the check box is if it doesn't apply, you don't have to do anything.

But to have that information and to have that information be something that can be transferred and interoperable helps the next setting sort of gear up to provide care for that individual based on what their needs are. So, I just wanted to touch on that.

And the PHQ, I just...for just a second if you'll let me dive in... talk about standardized assessment, PHQ-9 is asked in physicians' offices. It's standardized for a reason. It's a screening mechanism. Mental health and mood and how well is someone even getting their rest, those are critical things to know. And nurses, they assess a patient every shift in some way or another. So, it is really incumbent upon us to have a culture of safety and to have a rapport with our patients and residents. To be able to have a conversation, not just about depression, but about many things that you have to assess. So, hopefully the culture of the provider is such that they have the time and the tools that they need to perform those assessments. So, thanks.

Maria Edelen: Let's take a question from the phone. Do we have any questions from the phone?

Male Operator: We have a lot of questions. The first one is from Maureen Ledempser. Please go ahead.

Maureen Ledempser: I'm from an IRF in San Diego, California and I'm actually one of the nurses who was in the trenches doing the interviews with the patients. My question is about the PHQ and PROMIS. So, are you planning to utilize both or one of these?

Maria Edelen: We tested both of them and the plans for the future are still being debated.

Maureen Ledempser: If I have a vote...well, number 1, this is a lot of questions related to depression for patients and anxiety. It was the toughest part of the interview to get through with them and they actually would see patients who started out happy be brought down by the time it was

over. So, my recommendation would be to just use one and my personal preference would be the PHQ. So, I just thought I'd throw that out there.

Maria Edelen: Okay. Good to hear. Thank you. Something from the room?

Chad: I just have a question about the time that you're reflecting. I'm Chad from American Society of Consultant Pharmacists. I think med rec said it was done in 3.2 minutes. But what I guess my question is, is that 3.2 minutes the time that it takes the MDS nurse to fill in the boxes of the MDS? Or is it the time that it took to create the data that then the MDS nurse looks at to plug in to the MDS? Because you can't do a medication reconciliation in 3.2 minutes. You can document if you have the reports from multiple sources in the MDS, but I'm confused as to what the time actually means.

And it seems like it's a question that we're all kind of asking: can you do a PHQ depression scale in whatever the minutes were? Because it didn't feel like you're reflecting the time it takes to accomplish the task. It feels like you're reflecting the time it takes to actually document after the task has been completed.

Maria Edelen: These times are based on...yeah. That's a good question. So, these times are based on the actual how long did it take to go through this interview. So, for the PHQ, they write down the time when they start, they ask all the questions, and then they write down the time at the end.

The medication...well, we'll get to the med rec when we get there, and I can talk about that, but it's important to note that the data element for medication reconciliation is not actually performing medication reconciliation. It's a check on that process. And so, it's not as involved as going through a thorough medication reconciliation. What they're doing is saying did this happen? Did this happen? Did that happen? And so, they're just looking in the source material and determining...and it did in fact take on average the amount of time that we are reporting. Do we have another one from the phone?

Male Operator: Yes, we have one from Wendy Bunting. Please go ahead.

Wendy Bunting: Hello. My name is Wendy. I'm at an IRF and I'm concerned about the validity of many of these questionnaire-based assessments for our patients that have cognitive or communication

impairments. And I say that...I understand that patients with severe impairments would be identified, and you've got the non-communicative assessment for those. But many people have mild to moderate difficulty in understanding. Even just from a health literacy standpoint and so they may not be able to or they may be unwilling to admit they don't understand. And I'm worried those patients are falling through the cracks and we may be underestimating their needs and therefore not truly assessing the burden of care or where they need to be taken care of in the next setting.

Maria Edelen: I was going to say...that's really interesting to hear and I was wondering if you have particular experience with that because what we tried to do is separate out those who are really unwilling or unable to communicate. But we always are trying to get as much directly from the patient as possible. And so, what has been your experience with your patient?

Wendy Bunting: So, I am a speech language pathologist, so that is definitely my background. And so, what you'll often find, especially when you're just meeting someone and you're interviewing them, is they will tell you what they think you want to hear. So, there's a lot of nodding, there's a lot of yeses to paint themselves in a more positive light rather than to say, "I don't understand what that word means." They won't ask for you to express the question in a different way. They'll just be uncomfortable because they're in a situation not fully understanding and so, they just want to end the interview and rush it along.

So, for those people, especially because depression and anxiety are questionnaire-based, I'm concerned that they will be uncomfortable in the situation and just give what they think to be the positive answers because they are uncomfortable. And the length of time...often patients with mild cognitive deficit or communication deficits will not be able to tolerate a 20-minute interview without a decline in their ability to understand. They will basically be fatigued and confused.

I wish that there was a way that we could adjust...if cognitive or communication impairment was identified, that we could adjust their score or somehow to make note that we may be inflating their functional or independent ability.

Maria Edelen: I'm at a loss for words. This is definitely something that we need to think about and learn more about. We didn't hear from the IRF

assessors in the field very much about this issue. I think I might like to go back and see if we can delve into this a little bit more and learn more about how prevalent it is and how big of a concern it is. I think it's possible that you could do a companion assessment that's a staff assessment. I also want to say that the...we tested a lot of interview items and I don't know that any standardized assessment would ever have that much interview questions in a row. So, the fatigue issue I think would be mitigated by just a reduction in volume in questions.

Wendy Bunting: Just one little follow-up piece. And thank you for taking that feedback and being willing to ponder that further. Health literacy though is one thing that I'm wondering if it was taken into consideration or if it was taken into consideration, for how the instructions are scripted because the average literacy level in this country is fourth grade and that's just for reading. And then you throw in jargon and more medical terms and I'm concerned that people won't really comprehend what we're trying to get at when we asked them direct questions. Thank you very much.

Maria Edelen: Sure. Sure. Should I comment on that? I mean, as far as the health literacy, from my perspective I think it's a really important issue. To the extent possible, we didn't make up any new questions. So, it's either a variation on a question that's already being assessed in one or more settings or the PROMIS items which are geared to a really low literacy level. So, this is taken into consideration.

And we also conducted a series of cognitive interviews for any assessment items that were never introduced into a post-acute care setting before we went into the field with them. So, it was something that we really considered and made sure the instructions were conveyed in a way that was understandable to, at least a large majority of the patients and residents for whom it was intended. Okay. Another from the room?

Cynthia Morton: This is Cynthia Morton from NASL. I'm not exactly sure of my question, but I'll kind of blunder through it a little bit. Back on the continence, you are talking about the fact that there was lower reliability for the chart review when it came to continence. And it seemed to have a lower kappa. It just seems to me that that would be a little bit more cut and dried. And so, I was wondering if you would just talk again a little bit more about your experience for the assessors experience there. And were you comparing what was in the chart versus what was found from may be an interview, or were you comparing setting to setting?

Maria Edelen: So, all of the interrater reliability is based on a subset of patients and residents for whom the assessment was completed by the staff assessor and by the research nurse. So, they both went in and looked at the chart independently and said, "Okay, is this person using a device? Was the device placed in the current setting? What is the reason for the placement of the device?" And so, actually the interview items...it's a lot easier to write down the same answer because in the interviews, they didn't independently interview the patient or resident. They went in together and both wrote down what they heard the patient or resident say.

So, for the chart review there's a lot more room for error. And the fact that they're not abysmal, I think is amazing. I mean it says something for the chart. And also, the ability of assessors to find the information. So, they were saying, "It's hard to find. It's not always in the same place." And then of course, the staff are more familiar with the way things are charted, whereas the research nurses are not quite as familiar. And so, that's where we are getting less than perfect reliability. It's because they are both looking, and they are both looking at the same chart, but they're not always finding information.

Cynthia Morton: That's very interesting.

Maria Edelen: Yeah. Another one from the phone?

Male Operator: Yes. The next question is from Amanda Dawson. Please go ahead.

Amanda Dawson: Yes. Hi. I'm representing an LTCH system and I have three questions. One of them is, what you are basing your interrater reliability statistics on in terms of what level of agreement you're looking for. If it's actually at the item level, if you're using the scales, or if it is some kind of total or maybe a clinical interpretation off of those different measures. I can tell you that we've been doing a big interrater reliability study on CAM ICU between our nurses and an expert CAM ICU rater who has to evaluate the same patient within one hour. And we've been running that for...I don't know... its been about four months now across 100 sites on a weekly basis. And we don't get the same kind of interrater reliability that you're getting.

But then my second question is, when you say there's very little missing data, I want to understand if that includes UTAs. I want to say something in relation to the SLP that was just speaking a

moment ago from IRF. We'd actually been working with Dale Needen's group who has an Outcomes After Critical Illness in Surgery group, OASIS group, defining outcome measures for ICU survivors and specifically in areas like cognition, depression, anxiety, and so forth. We've been using the OASIS Delphi Panel approved measures that the Hospital Anxiety and Depression Scale as well as the EUROQUAL and we've been seeing very similar things to again, what the SLP was saying earlier, which is that we have a lot of heterogeneity in our patient group. So, I am very concerned that there are certain ones that are going to be rated as UTAs simply because they have cognitive deficits, not necessarily communication deficits.

And a lot of it had to do with them not being able to follow those scales. Like the scales that you are using are actually quite complicated for our patients. So, we've been having a lot of trouble applying the Hospital Anxiety Depression Scale especially. And indeed, patients have been saying that it actually makes them kind of depressed and anxious to hear those questions. So, I feel that there would be incredible heterogeneity. And who could actually answer that and answer it appropriately. Indeed, many of them do give answers that are in no way reflective of what their actual capabilities are.

And then my last question had to do with the way in which you're changing these scales. Like you are kind of making short forms on the fly and it could really change the way that you interpret the scale so that whether they are clinically relevant and useful, if you no longer have the same kind of scoring guidelines as you would for the full scale. And then also, different questions contribute differently to the different latency factors behind those scales. So, kind of modifying the reference like the time period or removing items is not trivial. I would think you'd kind of have to go back to the study...the authors of these scales and that usually takes a lot of evaluation. Anyway, those are my questions. Thank you.

Maria Edelen: Okay. Thanks. That's a lot. Let me start with the last question first and just...I want to reassure you. I'm a psychometrician and I'm not doing anything lightly in terms of changing scales, or selecting items, or creating scores, or anything. And we...so for the PROMIS Depression and PROMIS Anxiety we are working hand in hand with scientists at Northwestern who are the stewards of the PROMIS item banks. And if you're aware of the item banks, although it's not trivial to select a subset of items from the banks and assume that they reflect the construct equally well, that

is the purpose of the item banks. They were built for that purpose. And so, it's possible to determine the validity and reliability and coverage of the items that were selected from the bank and confirm that it is sufficient. And the folks from Northwestern are doing those analyses and have been doing those analyses to ensure that the subset of items that we included in our test sufficiently covered the construct. So, I hope that's reassuring. If you want to check with me further about that, I'm happy to answer some additional questions over email. But I just want to say that there's nothing...we're not being casual about this at all.

Similarly, with the PHQ, the screening out that we adopted for the beta test is one that has been adopted previously by others. Some people just use the PHQ-2. And this tries to be something that straddles that line between maybe the PHQ-2 isn't enough, but maybe, in many cases, the PHQ-9 is not critically necessary. And so, our test tried to straddle that line. It's not something that we made up, it's something that is done, and we are currently consulting with Kroenke, who was the developer of the PHQ as well as Pfizer to confirm that this 2-9 version is something that they would support and feel comfortable with. But we know that it is used in other places.

So again, a really good point and a really important thing to recognize, but something that we sort of had on our radar as well.

There was something about the reliability. And I wasn't entirely following your first question about the test that –

Amanda Dawson: The interrater reliability. What level are you determining that?

Maria Edelen: Yeah. So, it sounds like you're doing test, retest reliability which is slightly different than interrater reliability.

Amanda Dawson: No. No. We do interrater reliability. No. We calculate **[inaudible] [02:16:25]** self and so forth. We do a lot of that.

Maria Edelen: Okay.

Amanda Dawson: So, no, that's not what I mean.

Maria Edelen: Okay.

Amanda Dawson: But it depends on where you're...what is it? Is it the item level? Is it the actual scale? Are they actually trying to match the same

scale because you think the 100% agreement as well as the kappa? Or is it the total? Or is it kind of an interpretation of like normal, not normal?

Maria Edelen: So, the kappas are... we're giving ranges of kappa that are for each item.

Amanda Dawson: Right. Okay.

Maria Edelen: So, this item, I felt lonely, what was the agreement? Never, sometimes, often, always –

Amanda Dawson: Okay. On the item level. On the rating. Okay.

Maria Edelen: On the item level. There are very few data elements in our assessment that have a relevant scale. The majority of these are single items and are not scaled up. So, it's really just the anxiety and depression that are more amenable to scaling. So, a lot of your experience I think...it sounds like from the scale perspective and a lot of the work that we have here is more at the individual item level.

Amanda Dawson: Right.

Maria Edelen: We're really just saying did the research nurse and the facility staff record the same response to this same query? This single item.

Amanda Dawson: Right. Right. Well, I'm just curious because percent agreement is so prone to inflation due to the lack of correction for chance. So, if it's kind of a yes or no question, that inflation is aggravated. So, anyway, I'm just curious. These are high levels for some of these measures compared to what I've seen in the literature and what we have measured.

Maria Edelen: Yup. Duly noted. I think we have some language sort of about the considerations for kappa and percent agreement and...in the fine print that we don't have time to really get into here. But again, if you want to reach out to me via email, I'd be happy to continue with you. A middle question that I didn't –

Amanda Dawson: It's the heterogeneity. We are applying on the haves and the **euroqualitory** patients right now and we're finding there's incredible heterogeneity. A lot of the patients...a very significant proportion are unable to do the hat because the scale is too difficult for them. It's only got maybe five items, but they can't hold onto

that scale and answer those questions and they can't get through even a handful of items before their attention has been pulled or they are weary in relation to what was being said earlier. We have been presenting on this. We have data on this now. So, I don't...what I'm curious about is whether you've been looking at...if UTAs...Unable to Assess... is actually counted as missing or if it's considered a data point? And if it's a data point, if you were looking for subgroups and heterogeneity in patient populations. I know you said there weren't many differences in settings, but frankly the acuity is quite different in the settings. And I would think you would see a lot of UTAs in the long-term care...long-term acute care setting.

Maria Edelen: So, I mean if...it's really interesting to hear about your experience. I can only say that when we discussed our selection of the providers and our selection of the settings and our selection of enrolling their patients, we tried to get all the patients who were admitted during the field period. And we attempted to get a reasonably representative set of settings...each type of setting...and we didn't hear a lot about this. So, we've heard from you and we've heard it from someone else on the phone. It's not a non-issue. And it's certainly something that we need to circle back and consider.

But I can say we, not for lack of trying or not for lack of querying, did not hear this from the field in our beta test. So, we're going to circle back and look into it. But I think there might be something a little bit different about the specific populations that you're dealing with.

Amanda Dawson: Yeah. Similar to what was being said earlier I think there might be specific patient populations you're not going to be capturing because of these cognitive issues that they have. And a lot of our patients are emerging from minimally conscious states and they have significant awareness and cognition issues that are not just communication based. Is it the case that you called UTA missing? Or is missing just a blank? That would help me understand a little bit.

Maria Edelen: I'm sorry, I don't know what UTA is.

Amanda Dawson: Sorry. 'Unable to Assess'. You had an 'Unable to Assess' category for measures.

Maria Edelen: Yes. Yes. So... oh, okay. So, one thing we could do actually is look

at patient characteristics for those who...I'm looking at my data analyst here.

Amanda Dawson: Yeah, because you're saying you had very little missing data. I guess that means that these measures are appropriate for these patients because people aren't skipping them. But if a lot of people are putting 'Unable to Assess', that gives you another story about whether these are appropriate or not.

Maria Edelen: I thought the 9s were counted as missing.

Amanda Dawson: That's what I was wondering.

Maria Edelen: Yes. We counted them as missing.

Amanda Dawson: Missing. That's what I was wondering. Okay.

Maria Edelen: Yeah. I think the not applicable...well, if it was not applicable, it was not counted as missing. But the 'Unable to Assess' was counted as missing. So, we had pretty low rates of 'Unable to Assess'.

Amanda Dawson: Yeah. Because many of our patients...we have to use coma scale. Like they're not...you can't...I guess you put those in non-communicative though, but that's fine.

Maria Edelen: Yeah. Right. We'll have some more on that.

Amanda Dawson: Thank you very much.

Maria Edelen: One more question from the room and then we'll take a break. One more from the room.

Cynthia Morton: Cynthia Morton again from NASL. I almost hate to raise it after the dialog we just had, but...and I'm not a PROMIS expert by any means. Are the PROMIS questions, are they appropriate for the post-acute care setting patients? I mean I know SNF the best and I just don't know that you can...were you trying...did you use questions that you know you could get an accurate answer from the patient? Because the information is coming directly from them?

Maria Edelen: Right.

Cynthia Morton: What if they have cognitive problems and their answer wasn't going to be accurate? Did you look into that?

Maria Edelen: Well, I have to...cognitive issues aside, I can answer that question. So, one thing to know about PROMIS is that...so for depression there's this big bank of questions about depression and depressive symptoms. And what we did to try to select the subset of questions that were most relevant for post-acute care was we sent out the entire list of questions to a large group of stakeholders and asked them to rank their relevance to post-acute care. So, you remember it. Excellent. So, we took your input and combined it with others' input and we took the subset of questions that were seen as most relevant for post-acute care.

Yeah. Okay. All right. Now we're at 2:40pm. I think we can take another quick break and get started back up in 10 minutes. Thank you. **[Recording off]**

Maria Edelen: Okay, everybody on the phone, we're just getting settled back down. We're about to finish up. Okay, great. So, we have one more set of results...or two more actually. We've got the 'Non-communicative', but we also have, for the communicative assessments the data elements that fell into other clinical categories. So, these include the data element reflecting 'Assessment of Medication Reconciliation'. This was field tested in Alpha 1 and also field tested in Alpha 2 and also included in our second Public Comment period.

There's a little bit of this in some of the assessments. As you can see, the drug regimen review, which is a recently standardized measure, is in all four assessments. The MDS also lists or records whether certain medication classes are being taken by the resident. We also have 'Care Preferences', which included decision-making preferences, whether or not there's a designated healthcare agent. And these data element items were tested in Alpha 1 and in Alpha 2 in various versions, and we also solicited public comment on them in our second Public Comment period. And currently, the MDS includes a question about the patient's or resident's preferences for involvement with care decisions.

And finally, we tested the 10-item PROMIS Global Health which was included in Public Comment 2 and was discussed by the second convening of our technical expert panel. And since this is a standard 10-item questionnaire, it wasn't one for which we asked stakeholder input on which items would be most relevant because we decided to just go with this 10 item version from PROMIS.

But we did include the two different versions. So, one version is

as it is in the PROMIS library. And the other version asked about patient's and resident's experiences in the past three days.

So, now I'm on slide 74. We're going to talk about the 'Medication Reconciliation' data elements. So, this is on pages 50 to 53 of the communicative protocol. And this is how this appeared on the tablet. So, there were six medication classes that were queried. Is the patient or resident currently taking any medications in any of the following classes? And so, first, if they said yes, then there was a follow-up. The next items would show up on the tablet and ask to be answered. If they weren't taking any medications in any of the classes at all, they would be done with the interview. If they were taking only one or two then the follow-up questions are only asked obviously for those one or two questions.

So, first they're asking whether or not the medication classes are being taken. Then whether or not there was an indication for the classes that are being taken, whether there were discrepancies recorded involving any of the medications, whether or not the discrepancies were the patient's or resident's discrepancies addressed by involving the patient and their family and the formal caregiver? Where they addressed within a certain amount of time? And then finally, was the reconciled medication list communicated? And to whom?

So, this basically, as I said earlier, wasn't an actual thorough conduct of the medication reconciliation. It was more documentation that the appropriate steps that should have occurred or in an ideal world should occur, that they in fact did occur. So, this took 3.2 minutes overall to complete and it obviously took longer for patients and residents who had medications in many of the classes. So, if they had medications in more classes, you had to search more documentation to find out the follow-up questions. If they had only one or two medications, the section went quicker.

The percent agreement for this was pretty high...79-96%. The kappas...we couldn't compute kappas for all of them because as you cut down, the sample sizes were getting smaller and smaller and some of these medications are not that prevalent. And so, similar to some of the other data elements, we couldn't get kappas across the board. But where we did compute kappas they were ranging from 0.4 to 0.89 and they were highest for just the notation that the class was taken. There's a lot of agreement on 'is the patient or resident taking any of these medications'. And also, the

indication one had a pretty high kappa.

In terms of feedback, it was considered to have very strong clinical utility in both the assessor survey and according to the researcher's teleconference. The assessor's survey responses indicated that they found this to be particularly useful in home health. And also, both the research nurses and the facility staff saw this as being very highly useful for transfers, especially just to make sure that the patients are safe as they transfer from one place to another.

Despite all of the work that we've done to reduce the burden from Alpha 1 to Alpha 2 to beta, it's still seen as having the highest assessment burden. It is what it is. It took several pieces of documentation that are required and, as I said, a patient who is on more medications is going to have a higher burden. So, its still somewhat relatively higher burden than a lot of the other data elements that were tested in our beta test.

It was also somewhat challenging to understand the discrepancies. This was something that came from the research nurses. I think, not just to understand the discrepancies, but to also understand whether what they're seeing in the documentation is in fact the discrepancy. And so, it was more of how do we take what we're seeing in the chart and directly apply it to what we're being asked to record in this assessment. And so, I think that's something that could be mitigated to some extent with more training. And also, with sort of more routine use then the documentation might be a little more on point because then you know you have to write this down in some other place.

And then there was also the documentation on whether or not there was follow up and whether or not there was communication...a lot of times they just couldn't find it so, it was hard to say, no it didn't occur, or yes it did occur. It was more sort of I don't see this documented anywhere so it's hard for me to answer that question.

So, despite all of the changes that were made, there were still aspects of this protocol or this set of data elements that were somewhat challenging. But the clinical utility and the safety aspect of it were also really strongly supported.

For 'Care Preferences'...this is on page 40. There were I think, two questions...three questions that we asked. Two of them were asked of the patient or resident. They were asked about the extent to which they wanted their family and friends to be involved in

their care decisions. And then also, how much did they want to be involved in their decision-making. And then a third question was about whether or not there was evidence of advanced care directives in the chart.

So, this took about a minute and a half to complete overall and there were very few setting differences. The percent agreement for this data element was pretty high...83-99%. And the kappas were reasonable to excellent. So, the existence of the documentation of the healthcare agent was the one aspect of the data element that had the lowest percent agreement and kappa. It was just hard to find it in the charts and sometimes some people...one rater thought they found it and one rater didn't. There was a lot of discrepancy on that one.

This is seen as having very high clinical relevance, particularly during care transitions and have the low assessment burden. This healthcare agent question, it really tripped people up. I'm not sure why, but that was the question that seemed to have the highest burden. The other ones were interview items, so they weren't really burdensome to collect. And as we said earlier, the formal documentation of the healthcare agent was rarely present. Especially in the home health setting. So, I think it was a little bit frustrating for the assessors because they felt like they just couldn't find the information.

Okay, so finally we have 'Global Health'. This is a standard 10-item general quality of life inventory that was developed by the PROMIS Group to reflect overall quality of life. And you can also derive mental health quality of life and physical health quality of life summary scores for this. And actually, in the larger beta report, the PROMIS Group is doing some analyses separately for physical and mental health quality of life. It's actually pretty interesting. This is on pages 9-12. It took three and a half minutes to complete the data element overall, and there weren't a lot of setting differences in the time to complete.

There was a very high reliability as with all of the interview items. And as far as assessor feedback, it was somewhat to moderately clinically useful. We didn't get a lot of assessor feedback on this in terms of support. We did hear some challenges and concerns primarily that there was just a question as to whether all these questions were really relevant for post-acute care setting. And this is one where, as I said, we didn't take that first step of giving the stakeholders a large item bank and asking them to select the most

relevant questions. This is a sort of canned inventory. We took the 10 items and we administered them. We did do cognitive testing and the Global Health 10 was also previously administered in a home health setting. And so, we knew that it wasn't completely irrelevant, and it was reasonable to ask this set of questions. But still, people did find that a lot of their patients and residents didn't feel like the questions really pertained to them in their current setting.

They also found that it was difficult to report average pain particularly because pain varies so much and so they didn't really...they just had a hard time answering that particular question.

Okay, so now we're finally on to the last bit which is the 'Non-Communicative Data Elements'. So, for this we had staff assessments of mental status, mood, and pain. So, these are in the other...I'm not going to...I don't have the non-communicative...it's on page...the staff assessment of mental status is on page 3 of the non-communicative assessment.

This is similar if not identical to what is currently assessed in the MDS. So, for a patient or resident who is unable to complete the BIMS, then the staff assessor goes on to complete the staff assessment of mental status. And so, they indicate whether short term memory is okay, long-term memory is okay, whether the person has recall ability, orientation to the current season, location of the room, staff names and faces, and there's also a question about cognitive skills for daily decision-making.

This took 2.6 minutes to complete overall for the entire set of questions and had really high reliability actually. The kappas ranged from 0.74 to 0.94 and the percent agreement was quite high.

The staff assessment of mood is the PHQ-9 observational version. It took three and a half minutes overall to complete. And this one also had really high reliability...0.91 to 0.98 kappas. It's on pages 6 to 9 of the communicative protocol. This one actually tracks the same PHQ symptoms as the PHQ-9 that is conducted via interview. But the assessor is meant to estimate based on their observation whether or not these symptoms were expressed over the past two weeks.

And finally, we have the 'Staff Assessment of Pain'. And this one took just under two and a half minutes to complete the entire. It's

on pages four and five. And this also had pretty high reliability. The percent agreement was quite high and the kappas ranged from 0.81 to 0.90.

So, the feedback on the 'Non-communicative Assessments' was not a lot. Just in general, there wasn't a lot of discussion about these assessments. They were rated as moderately clinically useful overall in the staff assessment. So, they weren't...we didn't get strong support for them, but we also didn't see them as very problematic. They found it slightly more difficult to collect and a little bit more burdensome of course. I mean it's hard to determine these types of things with a patient who's not communicative.

And they also felt like if a patient is truly not communicative, their orientation to whether or not they know my name and face isn't even relevant. And so, there's sort of this fine line as to at what point can this set of questions even be ascertained for various patients and residents.

And this last point, I guess it could be maybe getting on some of what we were discussing earlier. We have the sort of borderline patients and residents who some might say they are communicative, and others might say they are not communicative because they're not cognitively able to really meaningfully get the scale and all that. And so, there was a little bit of confusion about whether this was the protocol for them or whether they should be trying to do the interview. But that's another thing that came up.

And now, I just sped through that. I am 10 minutes ahead of myself. We're supposed to be wrapping up. And then we'll end with some questions. Do you wanna talk about this last bit? Or I can if you want. Oh, wrong slide. (To Stacy: Do you want to discuss that or would you want me to? Okay.)

So, just as a final solicitation of your thoughts and feelings about all of this, we're collecting input on all of the standardized patient assessment data elements that were tested in beta. And all of what we tested was presented in this forum. And we're interested in your input, so please send us any input that you have to the e-mail address that you see on the screen. And the comments received by the close of business on January 15 will be officially reviewed and summarized and a verbatim comments summary report will then subsequently be posted on the CMS website.

I just want to be clear that we're not going to be responding to the

input. We're taking it in, we're going to read it, and we're going to synthesize it. We really want to hear from you and we want to know what you think. But what we're going to do is take it in and put it back out for everybody to see.

So, we have now some extra time and time for some final questions if anybody has any. Yes?

Cynthia Morton: Cynthia Morton with NASL. Back on that slide where you are talking about the guiding principles for evaluating the candidate spade, one of the principles is feasibility for post-acute care the potential to make these data elements interoperable across settings. Did you get into that at all in this study? I'm kind of putting my IT hat on. I know Liz is here.

Maria Edelen: Do you want to talk about that?

Stacy Mandl: We definitely have taken that into consideration. So, in fact, I'm not sure if you're aware, but you probably are of the CMS Data Element Library we have back here for that purpose. So, is the information meaningful? Is it something that would be useful to the next setting? To have, when the person arrives, assuming the data would be intraoperatively exchanged, so, yes. I think the answer is yes. That was part of sort of the Gestalt of getting consensus and working with the staff and the input that was reflected. So, yes. In the back?

Beth Conner: So, all of the assessment items will be assigned link codes. And when we can get LOINC and SnowMed codes we'll get other HIT codes and they'll all be placed in the data element library. So, then IT vendors and developers can then go in, grab those codes, and use them when they are creating software to help place the data into EHRs and in an effort to also support the exchange the information.

Stacy Mandl: Do you want to talk about tomorrow at your round table?

Beth Conner: Yup. So tomorrow, MITRE, one of the CMS contractors is holding an information gathering session with various IT vendors and providers. I think Donna is going to be joining from Maisel. And just to kind of get feedback on what the next steps are for the data element library in terms of interoperability and how we can make it more usable and get some feedback on what next steps should be for CMS. So, we're really looking forward to hearing the input.

And Liz and myself and Terry O'Malley have a presentation on the Dell and what's going on at ONC's Annual Meeting on Thursday at 3:45 PM if anybody wants to join.

Maria Edelen: Do we have any questions from the phone?

Male Operator: We do. We do have one from Samantha Colby. Please go ahead.

Samantha Colby: Yes. Hi. I had a question about the medication reconciliation and the timing of how long it takes to complete it. I notice on the slide that you reference the admission assessment page 50 to 53. Is that 3.2 minutes only for the admission assessment? Or is that supposed to also include the discharge assessment?

Maria Edelen: It's just one assessment. So, they did it twice. They did it once on admission and once on discharge. Did we do times for discharge? I'm looking at Anthony. So, all of our time estimates are based on the admission assessment.

Samantha Colby: That could be problematic in terms of accuracy, particularly on discharge. Currently on the LCDS the questions that we have require us to look back over the patient's entire stay. And the question is not limited to just these drug classes. The question is were clinically significant medication issues found? And were they addressed by midnight of the next calendar day? The time it takes to complete these, particularly for long-term acute-care hospitals who have patient stays of 30 days, 60 days, 90 days, a year to do that entire review, particularly on discharge, it cannot be done in three minutes. And so, if you're only evaluating this on admission, how are we to effectively implement this for discharge? And how are we to trust that this will be doable for discharge? Thank you.

Maria Edelen: Sure. I know Stacy wants to talk to this but let me just speak to a few of these things. So, first of all, we have timing data from the discharge assessment. And we can look at it. For the majority of data on this it's not... it's sort of not necessary because you're just asking a question, but I see your point about the medication reconciliation. But I want to also distinguish between the data elements that we tested for medication reconciliation, which is not the process of medication reconciliation itself, but more the existence of documentation that that occurred. So, that's one issue.

And the second issue is that this is not meant to...it's meant to complement the drug regimen review, which is what I think you're

talking about. And not to supplant it. So, I'm going to now give the rest to Stacy.

Stacy Mandl: And any one of my colleagues from CMS as well. So, the drug regimen review. So, the impact that required a quality measure specific to medication reconciliation. So, let's just take a step back for, not just LTCHs but for all provider types. So, medication reconciliation was recognized by Congress as being so important that they wrote it into law. And the measure that is in place right now looks at medication reconciliation and looking at the drug regimen review to make sure that discrepancies or issues were addressed. Because that is one of the leading causes of mortality in transitions in care.

So, I totally agree and completely understand what you're describing, but I think that the significance of having a quality measure that actually looks at whether the things that need to happen to protect patients are happening. And that's really the intent of the measure. And obviously, the intent of the law. So, I just wanted to touch on that.

And the other thing to touch on is that the measure was actually adapted from the home health setting where the length of stay can be even longer. So, I just wanted to sort of touch on that. And I understand. I am a nurse. But hopefully in your LTCHs and in the other providers since medication issues that can lead to harm and adverse events including death, are being addressed in a timely manner and the documentation reflects that.

Samantha Colby: No one is saying that medication reconciliation is not important. And that was certainly not my intention to convey that. Of course, it is critical. You mentioned that this was about transitions of care. And so, the current questions...it sounds like what you tested was looking specifically around admission and discharge versus what we have been asked to do now, which is for the entire length of stay. And yes, while home health may have longer length of stays then LTCHs, patients who are in home health are generally on fewer medications then patients who are in LTCHs. And so, the volume doesn't quite transfer over.

Maria Edelen: Okay. So, those are definitely important considerations. Let's go to a question in the room.

Male: Sorry. So, to address what she had mentioned about this being a way to mark the elements of medication reconciliation occurred, will we have access to the raw data from this study to be able to

assess whether those things have occurred? For example, in the MDS...I'm a long-term care pharmacist, so that's my perspective. We already have to put in all the medications. So, I could run a query of MDS data and tell you whether or not that patient at those different times was on a medication in any of those six drug classes. I certainly wouldn't need a nurse or an MDS coordinator to fill in those check boxes. So, we have an opportunity to take a look at the data that was reported and be able to compare it against the standard that we know is reliable and assess whether or not this tool to capture whether or not people are doing medication reconciliation is actually valid.

Maria Edelen: Yes. And we're actually doing...I think I mentioned earlier that we are getting assessment data. So, we're getting MDS data for all of the SNF residents who participated in beta. And we're getting IRF-PAI assessment data for all the IRF patients who participated in beta. And we're comparing where relevant the beta data element to its sort of corollary in the standard assessments. So, we will do that. We don't actually have that on our list, though. So, thank you for noting that. You can check to see how well our medication reconciliation class list matches up with the SNF and MDS list. For sure.

Male: Okay. But that data...you won't be...will you be showing us that data?

Maria Edelen: We can include that in our beta report. Yeah. I mean I don't know that we can...the ultimate plan is to share the data. That's after the contract period. The beta data will eventually be available but matching it up with the assessment data and all that...I mean that pretty much goes beyond the pale.

Male: Sure. Sure. I just want...I'm sorry for being...not having paid attention, but when the assessor was doing these functions, and they had a tablet, and you measured time from the time they started that particular module until the ended it, that's where you get the time data for these different elements?

Maria Edelen: Yes.

Male: So, to your point, what you said earlier about medication reconciliation is they were probably paths done outside of that and when the nurse sat down and said, "I'm going to fill out this module and use the tablet, they were accessing the chart and maybe other places that they have gathered data, and then filled out

that module. And it was that time that was captured?"

Maria Edelen: You want this to be longer than three minutes.

Male: Well, no. I agree that it probably is three minutes if you have all the information here –

Maria Edelen: No, well...

Male: – and you're filling it out here, it probably does take three minutes.

Maria Edelen: Yes. Yes, but they were instructed to record...they were trained to record the time that it took to do the whole thing.

Male: So, your self-reported time, not the computer capture.

Maria Edelen: The tablet...we have both sources and we compared the sources. So, they wrote it down and the tablet sort of recorded it once they entered. But I...and I honestly can't speak directly to what really happened in the field all over the country.

Male: Neither can we.

Maria Edelen: No. But I could...if you send me an email, I can ask some of the people that were a little closer to the operation that could tell me a little bit more about exactly how people were trained to record the time. But I know that the gathering of the information was meant to be reflected in the time estimate.

Male: Okay.

Maria Edelen: Whether it was consistently done so or not I can't confidently speak to, but I think we did a good job training and monitoring and I hope that that three minutes reflects them getting into the chart, finding the information, and checking the boxes. That's what it's supposed to reflect.

Male: But the information getting into the chart would not have been reflected.

Maria Edelen: No! Not into the chart.

Male: So, someone performing the med rec and putting it into the chart, that information is not reflected.

Maria Edelen: That's a completely different exercise.

Male: Okay.

Maria Edelen: We're just looking to see whether it's there.

Male: Gotcha. Okay.

Maria Edelen: Yeah. Okay.

Stacy Mandl: And I'll just remind you that medication reconciliations are I believe a requirement under the COP so, that should be happening anyway. This is really...if you think about the use of the data element to also drive clinical decision support and care planning and identifying drugs that are high-risk and is there an indication for their use and that sort of thing is really pretty pivotal. I just wanted to add that.

Maria Edelen: Do we have another question from the phone?

Male Operator: We do. It's from Deb Head. Please go ahead.

Deb Head: HI. This is Deb. And I'm coming from an IRF setting. I have a couple of items that I want to just briefly talk about. I do have concerns that this is a really small sampling. And especially when you're talking about if there's questions about the consistency of really implementing the tools. So, I just want to put that out there.

I also have some concerns with the BIMS and some of the cognitive test things. I know just to kind of piggyback on one of the other callers, when we see a patient that has a top score on the BIMS but then showing mild to moderate cognitive functioning, it's not necessarily capturing the full burden of care or the implications on that. And so again, looking at some of the issues with cognition I think need to be delved into a little bit deeper.

But my question is regarding when you look at the PHQ and some of the other cognitive type questions, is there a risk for the patient being asked those questions multiple times. As you said with such as a PHQ, that that can be asked in the physician office and part of your regular annual physical or appointment, but if a patient is in the acute-care hospital and then goes to a post-acute care setting or two or potentially three, what's the risk of having those questions asked multiple times in a short period of time? And does that become overly burdensome on the patient or on the consumer to

where it really isn't adding value, it's actually taking away value from the patient's care needs. So, would it be looked at...if the patient is been asked this assessment question within the last 30 days, you can bypass this or some consideration where there would be a time limit that the patient is not repeatedly asked the same questions. Thank you.

Maria Edelen: Yeah. Thanks. I mean...I guess...I'm not sure how to sort of evaluate or mitigate this sort of repeat assessment risk for something as important as depression. I mean it asks about the past two weeks really for clinical reasons. And so, if you went with somebody in the doctor's office and they asked you this two months ago so we're not going to ask you again, I mean that's not going to work. So, I don't want to really speak out of turn because I think that you're asking an important question, but I don't think it's...I mean it seems really important to stay abreast of a patient's mood. If they take a turn for the worse, it has all kinds of clinical implications and it seems reasonable to go ahead and ask these questions.

I don't know that somebody going through the care continuum is being peppered with questions in a way that becomes unreasonable, but if that were occurring, I would hope that we would hear about that.

Deb Head: Okay. Thank you.

Maria Edelen: Yeah. Thank you.

Tray Hillman: Tray Hillman from UDS again. Just a quick question. Is this the extent of information that RAND and CMS are willing to share in order for us to provide feedback and comments on? Or would it be possible for RAND or CMS to up supply like frequency tables or the responses to all of these items so that we can address the utility for case mix, see how frequently patients are being identified as having mood or anxiety or depression issues? Would that amount of data be available to inform our comments and provide the feedback that you're asking for to guide the ongoing research that's been done?

Maria Edelen: I'm not...we have frequencies. We didn't put them into this because it would just be too, too much to try to present. As far as whether we want to supplement your information for input, I'm going to defer to Stacy.

Stacy Mandl: I mean I think that **[inaudible] [03:03:31]**

Maria Edelen: Okay.

Stacy Mandl: We may not have it immediately, but I think that what you have...

Maria Edelen: Yeah. I don't think it would be a huge burden for us to make that information publicly available. I get a little concerned about...I guess I get a little concerned about it being misused or misinterpreted and so, I'm hesitant to just sort of put tables and tables of numbers out there without support and without commentary. And so, that's my reluctance.

In fact, we had some frequency tables here and part of it was about the time but part of it was like how do you give broad strokes frequencies? You can't do it. And we were really sort of straddling what's feasible and what's reasonable and what can we present that we feel will be backed up by sufficient information so that it's really clear what we're finding and what we're saying. So, I would like to consider it because I can see where you're coming from and I think I want to talk about it with my team a little bit more, and with CMS and see what are the pros and cons of moving forward with that. Because I can see it being really useful but also, I have some reluctance.

Male: Yeah. We'd really appreciate it because we've talked about the BIMS multiple times.

Maria; Sure.

Male: We're currently collecting that data in the Medicare environment in IRFs and other setting. And seeing whether the data you have in your beta test is consistent with the experiences we're seeing nationwide, would lend some guidance towards is this measure being collected the same way in your beta test the same way as it's currently being collected. Or to the upcoding comment before, are we seeing differences now between a beta test and what's currently being collected in the Medicare data.

Same thing with bladder and bowel. Bladder and bowel is being collected a different way currently within IRFs. Can we reconcile what we're seeing from a frequency of those patients experiencing accidents in the beta test versus those that are being recorded as having incontinent events today. So, we can kind of compare and contrast the frequencies of the data you're seeing versus what

we're experiencing in the data collection amongst provider sets. So, that's really what's kind of behind the question. We don't want to misinterpret the data at all, we just want to understand what are you seeing? How are you showing the utility for case mix? How many of our patients are showing cognitive impairments on the tools that you're testing versus those that are currently showing cognitive impairments other ways. Via the use of IC-10 codes, via the use of other factors that we are currently collecting in the assessment data that we have today. So, that's really what's behind it.

Maria Edelen: Right. And that's exactly what I'm talking about. I mean, to your credit, that is certainly of interest, but we have to also...it's not going to be completely comparable. I mean that's the whole point. It's not because the data were collected in the wrong way, but for one thing we screened out the non-communicative, so our rates of impairment are much, much lower. But once you take that into account and...So we did do a little exercise with the BIMS. We looked at the 2014 MDS data and we took out everybody who...no, we looked at expression and understanding...rates of expression or understanding for everybody...the full MDS data from 2014 and then we screened out...we excluded those who were unable to complete the BIMS and would have, for our purposes, been non-communicative. And the rates were very similar.

So, that... because we were a little bit worried about...we had rates of expression and understanding that were a little bit higher than we expected, but a lot of that had to do with our design. So, again, I feel like if you want to take that and start comparing, it's an uninformed comparison and it can lead to misconceptions. So, that's part of the concern.

Male: Yes. The only thing I can say...it would just be for the feedback and comment. And as you stated, you wouldn't be posting those specific feedback and comments, so again, you could probably restrict any of those comparisons that are being made inappropriately. So, it would just guide us from feedback and comments. Because all we have to really go on right now is the reliability metrics that you've shown and the qualitative feedback that you've enclosed within this presentation and then the previous presentation. So, from a data driven perspective we don't really have much to go on to provide you adequate feedback beyond our qualitative response about whether we feel these are clinically useful materials.

So, we want to provide as consistent feedback as possible to allow us to all go in a common direction toward standardized data that we know will be meaningful and useful. And without this information we may not be able to provide you with that feedback or understand how that may impact us moving forward. So, we appreciate anything you can do to provide that information.

Maria Edelen: Okay. We'll definitely talk about it. All right. Any more questions from the phone?

Male Operator: No questions at this time.

Maria Edelen: Okay. Anything from the room? Yes.

Female: I'm just curious how you handled patients that didn't speak English?

Maria Edelen: They were excluded. They weren't assessed. That was an enrollment criterion. Anything else? Okay. Stacy...wanna wrap it up?

Stacy Mandl: So, just to wrap up...first of all, thank you everyone for being here in the room. Great job Maria and team. And for those on the phone, this is hard to be on a phone for four hours, even with breaks. So, thank you for the input and the questions.

I want to just sort of close with not only just a moment of gratitude for everyone's input and time spent here, but also just to reflect back on this sort of journey that has included a tremendous amount of feedback from all of you, from the public at large, not just from providers but also from patients and driving sort of the direction that lead us to here to today. I can't even speak I'm so dehydrated. But has led us here today.

What was tested in the beta testing was really a combination of public comments, not only through public comment periods but also through the rule making process. Also, from all of the technical expert panels that came to meet together to inform RAND, from the provider and patient expertise areas and all the many open-door forum calls and that sort of venue. Really this hard work was a culmination of all of that input.

There's a lot to be said for the significance of having the information that clinicians need to care plan and to take into consideration for clinical decision support when someone is

departing that provider and arriving at the next provider and make sure that that information can be spoken in a consistent way.
So, I was an army nurse and we were able to get, using standardized information, to get patients from the deep zone where they were injured on the field to a MASH unit, to a hospital or a Medic in a neighboring country over here to the states using standardized information. It's powerful. And having something to start with is better than having nothing. And so, I just want to thank you. This is a huge list and the information has been profoundly helpful. So, have a great rest of the evening and thanks a lot.

Maria Edelen: Thanks everybody.

Male Operator: Ladies and gentlemen, that does conclude the conference call. We thank you for your participation and you can now disconnect your lines.

[End of Audio]

Duration: 193 minutes