National Quality Forum
Scientific Methods Panel
Fall 2021 Measure Evaluation Meeting
Wednesday, October 27, 2021

The Scientific Methods Panel met via Videoconference, at 3:00 p.m. EDT, David Nerenz and Christie Teigland, Co-Chairs, presiding.

Present:

David Nerenz, PhD, Henry Ford Health System; Co-Chair
Christie Teigland, PhD, Avalere Health; Co-Chair
J. Matt Austin, PhD, Johns Hopkins University Armstrong Institute for Patient Safety and Quality
John Bott, MBA, MSSW, The Alliance; The Leapfrog Group
Daniel Deutscher, PhD, MScPT, Maccabi Healthcare Services
Marybeth Farquhar, PhD, MSN, RN, American Urological Association
Jeffrey Geppert, EdM, JD, Battelle Memorial Institute
Laurent Glance, MD, University of Rochester School of Medicine and Dentistry
Sherrie Kaplan, PhD, MPH, University of California Irvine School of Medicine
Joseph Kunisch, PhD, RN-BC, CPHQ, Harris Health System
Paul Kurlansky, MD, Columbia University Center for Innovation and Outcomes Research
Zhenqiu Lin, PhD, Yale-New Haven Hospital

Jack Needleman, PhD, University of California Los Angeles School of Public Health
Sean O'Brien, PhD, Duke University Medical Center
Jennifer Perloff, PhD, Brandeis University
Patrick Romano, MD, MPH, FACP, FAAP, University of California-Davis Health
Sam Simon, PhD, Mathematica Policy Research
Alex Sox-Harris, PhD, MS, Stanford University
Ronald Walters, MD, MBA, MHA, MS, University of Texas MD Anderson Cancer Center
Terri Warholak, PhD, RPh, University of Arizona College of Pharmacy
Eric Weinhandl, PhD, MS, Fresenius Medical Care North America

NQF Staff:

Poonam Bal, MHSA, Director
Tricia Elliott, MBA, CPHQ, FNAHQ, Senior Managing Director
Hannah Ingber, MPH, Senior Analyst
Gabby Kyle-Lion, MPH, Coordinator
Elisa Munthali, MPH, Senior Consultant
Matthew Pickering, PharmD, Senior Director, Quality Measurement

Also Present:

Susannah Bernheim, MD, MS, Yale University Centers for Outcomes Research and Evaluation (CORE)
Jeph Herrin, PhD, Yale University CORE
Sriniketh Nagavarapu, PhD, Acumen LLC

Contents

Welcome and Review of Meeting Objectives 5

Advisory Discussion 9

Measure 3667 9

Re-vote Measure 3667 36

Measure 0689 37

Re-vote Measure 0689 51

Opportunity for Public Comment 52

Next Steps 53

Adjourn 55

Proceedings

(3:02 p.m.)

Welcome and Review of Meeting Objectives

Ms. Ms. Elliott: Excellent. Welcome back everyone to our Scientific Methods Panel Day 2. Just wanted to refresh everybody's memories of some of our housekeeping reminders. This is a WebEx meeting with audio and video capabilities.

Ideally, everybody's in and able to connect with the link and meeting number password, everything that was shared in the meeting insights. If anyone's having trouble with the audio, we do have an optional dial-in number. Please place yourself on mute when you are not speaking.

We encourage you to use the following features during the meeting. One is the chat box to message the NQF staff or the group. The chat box can be found in the lower right corner.

The raise hand feature is also available. And that can be found at the bottom of the screen. There's a reactions icon. A smiley face. If you click on that there's a raise hand feature. Or if you have the participant list open, you can hover near your name, and the raised hand feature shows up there as well.

We will conduct roll call in just a couple of minutes. And just a reminder, if you're experiencing any technical issues, please contact the NQF project team, either through the chat or the email at methodspanel@ qualityforum.org.

Next slide, please. So we'll do a quick roll call of the committee. David Nerenz?

Co-Chair Nerenz: Here, and no conflicts.

Ms. Elliott: Great, thank you. Christie Teigland?

Co-Chair Teigland: Hello, yes. I'm here.

Ms. Elliott: Great. Thanks, Christie. Matt Austin? Bihan Borah? John Bott?

Member Bott: Yeah, I'm here.

Ms. Elliott: Hi, John. Thank you. Daniel Deutscher?

Member Deutscher: Hello, I'm here.

Ms. Elliott: Welcome back. Thank you, David. Daniel, sorry. Lacey Fabian? Marybeth Farquhar?

Member Farquhar: Yes, I'm here.

Ms. Elliott: Thank you. Jeffrey Geppert?

Member Geppert: I'm here.

Ms. Elliott: Thank you. Larry Glance?

Member Glance: I'm here.

Ms. Elliott: Joseph Hyder? Sherrie Kaplan?

Member Kaplan: More here than I was yesterday.

Ms. Elliott: Great. Thanks, Sherrie. Joe Kunisch?

Member Kunisch: I'm here.

Ms. Elliott: Hi, Joe.

Member Kunisch: Hi.

Ms. Elliott: Paul Kurlansky?

Member Kurlansky: Hi, I'm here.

Ms. Elliott: Great. Thanks, Paul. Z.Q. Lin?

Member Lin: Yeah, I'm here.

Ms. Elliott: All right. Thank you. Jack Needleman? Gene Nuccio? Sean O'Brien?

Member O'Brien: I'm here.

Ms. Elliott: Sean, thank you. Jennifer Perloff? Patrick

Romano?

Member Romano: Here.

Ms. Elliott: Thank you. Sam Simon?

Member Simon: Good afternoon. I'm here.

Ms. Elliott: Great. Thanks, Sam. Alex Sox-Harris?

Member Sox-Harris: I'm here.

Ms. Elliott: Great. Thank you. Ron Walters?

Member Walters: Present.

Ms. Elliott: Great. Thank you. Terri Warholak?

Member Warholak: I'm here.

Ms. Elliott: Great. Thanks, Terri. Eric Weinhandl? Susan White? Okay, I'm going to circle back. I think Jen Perloff, are you there now?

Member Glance: Tricia, I wasn't sure if you heard me before. This is Larry Glance.

Ms. Elliott: Larry, great. Thank you. I did not hear you. Appreciate that. Okay, we got Jack. Jack Needleman, are you on now?

Member Needleman: Yes, I am.

Ms. Elliott: Great. Thank you.

Member Needleman: Still no conflicts.

Ms. Elliott: Okay. Sounds good. And I think Eric Weinhandl signed in?

Member Weinhandl: Yes, present. No conflicts.

Ms. Elliott: Thank you. Okay, there might be a few more that we check in with as we go. I think I'll have the team maybe message Jen. We think she's on and maybe double-muted or something along those lines.

Okay, she's having some audio troubles. So I think we are good. David or Christie, any opening remarks as we kick off day two?

Co-Chair Teigland: No, just, let's plow through our agenda here. See if we can come to conclusion on some of the things we left hanging yesterday. Thanks.

Ms. Elliott: Thank you.

Co-Chair Nerenz: Thanks to everyone for diligent work yesterday. Thanks for the team. We do have a little bit of restructuring in terms of how things set up this afternoon. I'll let the staff talk us through that. I think we have a couple things off the table now that might have been on, but I'll let our fearless staff leaders walk us through that.

Ms. Elliott: Great, thank you. So we will do a quick recap from yesterday, because it did impact, as David and Christie alluded to, our agenda for today. Hannah, would you mind sharing? We got through six measures yesterday. Could you share the results?

Ms. Ingber: Sure, happy to do so. Yes, as Tricia mentioned, we reviewed six measures yesterday. 0689 passed on reliability and did not pass on validity. 649e did not pass on reliability and consensus was not reached on validity.

3650e was discussed but we did not re-vote. It passes on reliability and does not pass on validity. 3652e was discussed and we did not re-vote on it. 3638 we discussed but did not re-vote on that measure.

That measure does not pass reliability and validity. And 3639 we also discussed but did not re-vote on that and the measure passes reliability and validity.

Ms. Elliott: Great, thank you. So our focus for today will be starting with the seventh measure that we

did not get to discuss yesterday.

We had also shared on an agenda that 3652e, that we discussed yesterday and we were going to continue some discussion today, the discussion around that particular measure landed with looking at adjusted versus non-adjusted data and some discussion around that. But the measure developer has withdrawn the measure at this point. So we're removing that item from the agenda today for discussion.

Our third measure that is, we had shared in a revised agenda last evening, was 0689. And we will discuss that measure just to confirm the committee's understanding of all the materials and items shared with regard to that measure, and make sure there's clarity there.

And then we will continue on with our regularly scheduled agenda, which is some of the CSAC update, the reliability thresholds, and some of the test maintenance, reliability, and validity testing. And we'll leave time at the end for public comments and next steps.

Member Needleman: Okay, did I hear you correctly that 3652e has been withdrawn?

Ms. Elliott: Yes, yes.

Member Needleman: Thank you.

Advisory Discussion

Measure 3667

Ms. Elliott: Yes. Okay. Any other questions before we move into our last measure, which will be a full discussion on 3667?

Okay, so with that, I'll hand things over to Matt Pickering, NQF Senior Director, to discuss the

measure.

Dr. Pickering: All right. Thank you, Tricia, and good to see you all again today. So I'll just be talking about 3667, and then I'll turn it over to Dr. Romano, who's our lead discussant for this measure.

So, just a quick summary. This measure did pass on reliability and validity, but was pulled for discussion. And so some of the elements you can kind of see listed on the SMP discussion portion of the slide there. This is located on Page 16 of the discussion guide. The developer is Yale CORE, and the steward being CMS.

And this measure is a new measure. It's a provider group level measure of days at home or community settings. And those community settings being that it's not an acute care, such as inpatient hospital or emergent care settings, or post-acute care settings, such as skilled nursing facilities. And it's looking at patients 18 years and older. It's Medicare fee-for-service beneficiaries with complex, chronic conditions who are aligned to participating provider groups.

The measure includes risk adjustment for differences in patient mix across provider groups, with an adjustment based on patient's risk of death. And additional adjustment that accounts for patient's risk of transitioning to a long-term nursing home is also applied to encourage home- and community-based care in alignment with CMS policy goals. So a higher risk-adjusted score indicates better performance with this measure.

So it is an outcome measure. The data source is claims. The level of analysis here is accountable care organization. So, as I mentioned, it's risk-adjusted, and the model includes 52 factors.

With the reliability testing, it did pass. And related to the reliability specifications, many SMP members

found the measure specifications confusing and occasionally arbitrary. This was especially true of the denominator statement which lacked the target population, conditions, setting, etc.

The SMP also expressed some concern that several concepts included in the submission were not documented as exclusions in the specification, and also questioned whether the developer considered other exclusions related to specific reasons for being accepted into an acute care or ED that might not indicate low quality of the accountable entity.

Related to the testing, it was conducted at the accountable entity level. The developer tested the measure using a split-half methodology using data from 2017 to 2018. They reported the interclass correlation coefficient of .8326 for the final days-at-home outcome measure between the two samples.

Some SMP members noted that the use of the split-half methodology is better suited for federal accountability programs with multiple years of data, particularly because the ACO assignment roles are adjusted annually.

Further, the reported ICC, or interclass correlation coefficient, may underestimate true reliability because scores are estimated using only half of each provider's data. So that's with reliability.

Just briefly with validity, again, it did pass the SMP, and the validity testing was conducted also at the accountable entity level. Its construct validity was assessed using Pearson correlations with six other ACO level measures, and those being representative quality conceptually related to the excess days of care for patients with complex, chronic diseases.

So, for the Pearson correlations, they ranged from -.5492 to .048, resulting in high inverse for unplanned admissions, which was expected. Moderate with other measures, and no correlation with the fault risk, another measure that was looked

at, and unexpected inverse correlation with patient experience, another one of those ACO measures that was used for correlation.

The developer did perform face validity testing, as well, of the measure specifications and the appropriateness of the quality assessment at the ACO level, with 19 of 21 TEP members responding for this Technical Expert Panel they convened.

For the statement posed to the TEP, does the days at home as specified -- does this measure, as specified, can be used to distinguish between better or worse performance of ACOs or provider groups? So, two members of this TEP indicated strongly agree, 15 indicated agree, and two indicated somewhat agree.

So there was a series of concerns from the SMP for validity testing. Some members noted there were three different risk models used to use and express concerns about lack of clarity about whether or how they are combined to get a single score on the validity on that approach.

One SMP member indicated is not clear why primary death data was not used, but a death risk model was used instead. SMP members had also raised concerns with the overall risk adjustment methodology testing and results, noting that the measure construction approach inappropriately lacked adjustment for many variables without theoretical or empirical justification as well as potential arbitrary measure weighting.

Some SMP members commented that many decisions regarding social risk factors appeared arbitrary and were not pervasive. So the C-statistic for the risk adjustment model was .738, for the mortality model, .760 for the nursing home transition, and deviance from R-squared was .170 for the days in care model. And Spearman's rank correlations were conducted and that was .346 for more days in care.

Some members really questioned whether there were meaningful differences in performance. And the developer also stated that differences of three days should not be considered trivial.

Members noted that it's not clear whether this equates to meaningful differences in quality of care manifested, for example, in differences in patient function and health-related quality of life.

And finally, the same SMP members that raised those concerns also noted that the developer did not appear to test between versus within ACO variants adjusted for risk factors. Another SMP member then noted a difference in three days could reflect variables not included in the risk adjustment model, or in residual effects are not fully adjusted.

So, with that, you'll see the questions up for SMP discussion listed on the slide here. And I'll turn it over to Dr. Romano to lead us through the discussion. Dr. Romano?

Member Romano: Great, thank you very much. Can everyone hear me?

Dr. Pickering: Yes.

Member Romano: Okay. So, I'm going to focus the discussion on validity issues. I think the vote on reliability was unanimous, as you can see, 11 to 0.

So, let's just discuss validity to keep the task simple. So, I'm going to highlight two or three issues that I'd like to hear some discussion and perhaps some engagement with the developers.

So, first of all, in terms of measure score validity, I want to be very clear that the developers did something that they weren't required to do, which was they offered us both evidence from face validity and from empirical validity of the measure score.

The face validity evidence strongly supported the endorsement of the measure. However, the

empirical validity testing did not. And, in essence, as Matt has summarized, they correlated their new metric with several other ACO metrics.

And those analyses are all problematic for one of two reasons. Either they're correlating with endogenous measures; in other words, measures that incorporate hospital admissions, and therefore, the concept of days at home is inherently correlated with hospital admissions and hospital readmissions. So, that makes it impossible to interpret several of the correlations that are in the expected direction.

As we've discussed before in this committee, we're particularly interested in process outcome correlations to see what are the pathways by which outcomes can be improved. And here, they found a counterintuitive effect. That is that the ACOs that appear to be offering patients more timely care, more timely appointments, actually had more excess days in care.

So, that is clearly sort of invalidating the measure from the empirical validity testing perspective. However, it's a new measure, and so we can rest on face validity. And we appreciate their honesty in presenting this evidence to us that was not entirely supported. So, that may be fixable, of course, as the measure gets into use.

Second is exclusion issues. So, the developers report no denominator exclusions whatsoever. However, it's not clear how they handle patients who are already long-term nursing facility residents at the beginning of the measurement period.

So, this is a very complex measure in which they're identifying days at home based on essentially subtracting out inpatient days, post-acute days, inpatient rehab facilities, psychiatric facilities, long-term care hospitals, emergency departments, and observation stays.

However, they don't subtract out the days that are

in long-term or residential nursing home care. So, this leads to a question about, well, what about those patients who are already in long-term residential nursing home care, and who, essentially, have no chance of having any days at home?

I would have expected to see that as a denominator exclusion. But apparently it wasn't. So, the developers might be able to respond further on that.

It might be that this is an ACO population in which nobody is in nursing homes, long-term nursing homes. I don't know. But the big issue that I think several of us on the subcommittee struggled with was about the way they put together three different measures.

And I'm going to try to briefly summarize, but of course, the developers will correct me if I got it wrong, but, basically, their primary measure here is a measure of days at home. And the principal way that they derive the days at home is by first estimating days in care. And this is done using a standard negative binomial regression approach with days at risk as an offset variable.

They estimate, in usual hierarchical regression model, they estimate the predicted days in care and the expected days in care for each patient. And they difference those two, so now they have an EDIC, or an excess days in care, for each individual patient.

Now, the problem with this, what you might call naive measure, I guess, is that some ACOs may have excess mortality. And essentially, I'm going to say colloquially that they're killing some people. And so it may be that if some ACOs have excess mortality, that that is reducing the number of days alive that drives the offset for the negative binomial regression.

And, therefore, the ACOs will essentially be getting credit for that excess mortality. And the developers

want to penalize the ACOs that have excess mortality, so they don't get any benefit, essentially, from having excess mortality.

So they come up with an exceedingly complex way to do this where they estimate a mortality model using a standard hierarchical logistic regression approach, predicted and expected mortality. And then they estimate a nursing home transition model in the same way for transition to a long-term nursing home situation. And then they essentially take the EDIC that came out of the negative binomial, and they adjust that EDIC based on the results at the hospital level -- or the ACO level, sorry, from the ACO model and the nursing home transition model.

The problem, for my sense, is that, first of all, this is very complex for people to explain and understand. But I think the formulas are simply wrong.

So, the notion is that -- let's say the SMR for an ACO is two. So that means that that ACO has twice as many deaths as were expected based on the characteristics of the patients at that facility or at that ACO.

So if the SMR is two, then they are multiplying that SMR by each patient's EDIC in that ACO. So they are doubling the EDIC estimate for each and every patient enrolled in that ACO based on the SMR of two.

So, for example, if an individual patient has an EDIC of 10 days, 10 excess days in care, the SMR for that ACO is two, they would adjust that to 20 excess days in care. But the problem is that the actual absolute mortality rates that we're dealing with are much smaller. They might be one percent, five percent. And so the actual impact of that twofold SMR on the days that are at risk to be in care is trivial, compared with the magnitude of penalty that they're imposing.

So they're taking a difference measure of the difference in the days in care and they're then applying a multiplier to it, a mortality multiplier that comes from another model.

And they're doing the same thing for nursing home transfers, but they've arbitrarily chosen a .5 adjustment, so they only impose half a penalty for nursing home transfers relative to deaths.

So I think a number of us struggled with the complexity of this approach, with the fact the formulas didn't seem to have a sound basis, and that, in their response, the developers point out that the absolute effect of this penalty scheme is small in the vast majority of cases. And I think we can accept that, but I'm still stuck on the fact that the formula is conceptually incorrect. I can't see a rationale for applying an SMR multiplier or a nursing home transfer multiplier to a difference of the excess days in care and incorporating that as a penalty scheme to penalize ACOs that have excess mortality.

If we're going to do that, it seems that we should be formally labeling this as a composite measure, rather than as a measure of days at home, or the converse of EDIC.

So, that is a summary. There was also some debate, I think, in our subcommittee about the social risk factors. Matt has summarized that. They did do a careful analysis. We can argue with their conclusions. This is a classic example of a measure that we would expect to be sensitive to social risk factors, because you have to have social support to live at home.

And so, personally, I was surprised that more of those social risk factors didn't seem to be statistically significant, but they did go through a process of assessing those variables for inclusion or omission. So, I'll stop there. Sorry, I'm taking up too much time.

Ms. Elliott: Thank you, Patrick. The -

Dr. Bernheim: Can we -

Ms. Elliott: Yes?

Dr. Bernheim: Just before, can I explain? I see other folks want to talk, but I'm worried that it may be really helpful for us to clarify a few things, just to get the facts out before we have a debate about all of the important things Patrick just said.

Ms. Elliott: Sure. I just want to defer to the Co-Chairs to make sure. So, David and Christie, Susannah from the measure developer, Yale, would like to address some of Patrick's comments. Because usually we have the subcommittee or sub-group continue conversation, but are you okay if we pause here and allow input?

Co-Chair Nerenz: Yes, I think that would make sense, and I think Susannah just mentioned that the focus is on clarifying, possibly correcting, perhaps.

I think that naturally could precede a broader discussion. Again, if a member of the sub-group really wishes to jump in now, we could hold for ten seconds. But I think that Susannah's proposed sequence would make sense.

Co-Chair Teigland: I'm fine with that.

Dr. Bernheim: Great. And David, I'll try to stay in clarification rather than defense mode. I'll do my best there.

First of all, thank you, Patrick. And I will say I have Jeph Herrin to thank when we get to a couple of the statistical things. I'm going to let him do the clarification there because I tripped over myself.

I mean, you raised a number of important things focusing on validity. First was related to the comparison measures. But a lot of what you were

focused on was the construction of the measure. I'd love to come back to talk a little bit about the comparison measures and the external validity, because I think that that would be more in defense mode. I think there's some important things to note about that. So we'll come back to the external validity.

But I just want to make sure the committee understands how and why the measure is constructed the way it is and what the impact of those extra adjustments are.

And then happy to hear -- no question it can be confusing. And Patrick is a remarkably experienced and smart measure developer, so apologies for confusing the committee. I think it is good for us to think about how to continue to explain this in a way that will make sense to folks.

It has been a bunch of effort to build a measure like this. Before you guys may have seen work by Ashish Jha's group for MedPAC on a similar measure concept that ran into some problems that are really hard to tackle.

So there's a real desire to have a measure like this in the world that really looks at whether we are successfully taking complex, chronic patients through the health system in a way that maximizes their time at home.

And it's really hard, because you have to deal with mortality, and you have to deal with what to do about transitions to nursing homes, because otherwise stays at home aren't that meaningful.So we constructed a measure of days in care, which is the inverse of the days at home, that is just easier to model. So that is a fully risk-adjusted measure, similar to many of our outcome measures that accounts for lots of comorbidities and other conditions. But in studying previous measures that have tried to address this, and talking to our TEP and talking to CMMI, we had to make some choices

about how to essentially balance the desire to keep people at home with the potential unintended consequences of a days at home measure.

And so that is the purpose of those other two models. They're applied at an individual patient level. The highest standard mortality rate is 1.25. We included in our additional information the impact that it ultimately has on the days at home.

The goal is to ensure that there is not an ACO that's judged successful at keeping people at home either because they have very high, higher than expected, mortality rates for their community.

And then I'm just going to speak for one word on what happens with nursing homes. So, Patrick asked a really important question. What happens if somebody during the performance year is in a long-term nursing setting? And what we decided with CMMI was that -- we're a little bit stuck with this somewhat arbitrary performance year, right? Measures always have a performance time period. And so, you know, if you start with a hospital admission, it's very clean, you know, when time zero is, but if you're looking at patients that are living in the community, we start with where they are in January.

And s0 the goal was to determine what someone's home was in January. So for patients who start the year -- and we did not want to -- other measures have just excluded anyone who's in a long-term nursing home, but we want to look at acute care utilization for those patients as well. So we wanted those patients in the measure. So your home is determined based on where you live January 1st. As you know, Medicare patients who are not in SNFs, but in long-term nursing, tend to stay there for a long time. So that really is where they live at the beginning of the year.

And then we look at whether they are home for the rest of the year. The additional adjustment is

because there's a lot of concern in our TEP around whether or not it's a good thing for these patients to transition to long-term nursing care and how to not count every single day you send in long-term nursing care. This would have been completely overwhelmed by those days in the long-term nursing care. But also to ensure that you didn't look as if you were keeping patients at home if they were all transitioning to long-term nursing care.

So again, that second adjustment, which is weighted less than the mortality adjustment, is to, essentially, balance potential unintended consequences. It's really meant to only impact the measure in the rare cases where there's an unexpectedly high rate of patients transitioning into long-term nursing care and making sure that that doesn't look like a wildly successful ACO.

So, that's the conceptual. I'm trying to stay in the clarification realm. I really am. Jeph, do you want to clarify any of the modeling pieces before the committee discusses?

Dr. Herrin: No, I have things to say about them. But no more clarification with respect to what Patrick said. So, no, we're fine.

Dr. Bernheim: Thanks.

Ms. Elliott: Thank you. So we'll circle back with sub-group comments. Jack Needleman, you have your hand raised?

Member Needleman: I do. Thank you. And Susannah, thank you. As usual with the Yale group, we've got a very detailed set of documents that present things very well. And there's clearly been a lot of work and thought that's gone into the measures.

That said, I think the validity of this measure is not demonstrated, and it is precisely over the weighting of the transitions to mortality and transition to

nursing home. One and .5, I heard Susannah say, well, we wanted to provide some penalty to ACOs, but we didn't want to -- you know, they are arbitrary numbers, and they are arbitrary weights. And we've got no documentation that those arbitrary choices were the right ones, yet they are an integral part of the specifications of the measure.

And, therefore, since they look arbitrary, I don't think the measure's validity has been demonstrated. And that's not a complaint about the work. I think that's inherently, you know, one of the inherent limitations of trying to go from excess days in care to the days at home measure that you're trying to construct here.

But I need more just -- I, frankly, need more justification of what the right weights are on the mortality measure and the right weight on the transition to nursing home measure than we have here. And I think more work is needed on this measure. And to think through what those weights are and think through what the consequences are for assessing ACO performance on keeping patients at home.

Ms. Elliott: Thank you, Jack. I don't see any other sub-group hands raised or any additional chat. So we can open it back up to the developers for a two- to three-minute initial response.

Dr. Bernheim: I'll first respond to Jack and see if Jeph wants to jump in as well.

I mean, I think, you know, the SMR is essentially not weighted, right? So, the decision-making that we had was what to do when it was clear we wanted to have some adjustments for the tendency to transition folks to nursing homes and how to balance that against the mortality.

So that's the first key thing. You know, I think there -- you know, we sometimes make non-empiric decisions in measures, right? I mean, we weight

things, we decide a meaningful difference is. We do this with a lot of input. I don't know that there's a perfect empiric way to do this.

It was clear from our TEP that it was important that if you were an ACO that had way more deaths than expected, that that was a bigger concern than if you were an ACO that had a higher tendency to transition people to nursing homes. And so it was important to the experts we worked with that there would be a down-weighting.

First, they're standardized; they're put on the same scale. And then it was important that there be a down-weighting of that nursing home transition. But I can't pretend that we could do some magic math to decide exactly how much it should be.

I think the important thing about that is that the fundamental measures, the days in care, these have actually showed relatively small impact. They're really meant to balance against unintended consequences of this measure; they are not the heart of the measure. And with the TEP's support and general direction, it's the kind of thing where, if you changed, it wouldn't change the results a lot. So we don't feel like the somewhat inherently arbitrary nature of it would impact the overall validity of the measure, because it is a small weighting, as opposed to the sort of heart of the measure.

But, Jeph, I don't know if you want to add anything. We can go back to the empirical validity if you want it. I know, Tricia, we went out of order.

(Simultaneous speaking.)

Member Romano: Could I add a response to Susannah? Sorry. I'm just going to challenge the statement that the one is not a weighting. That one is a weighting also on the SMR. I mean, again, the practical problem here -- and I look forward to Jeph's response -- but you have an offset in the

negative binomial regression model for the number of the days the patients survived in the performance year.

Essentially, you think that number may be too low at ACOs that have excess mortality. In other words, they should have had more days when they were exposed to the possibility of being in care.

But, again, the practical problem here is that if the absolute mortality, let's say, is one percent, and the SMR, at the extreme, is two. So you have an additional one percent of patients that lost half a year, on average. And so the actual magnitude of the impact of that on the offset term in the negative binomial regression is tiny compared with doubling the EDIC estimate.

So, when I see in Table 4 that the maximum adjustments are plus four days, plus 12 days, the minimums are minus 10 days, minus five days, I don't find that terribly reassuring because those are pretty substantial changes. And the choice of using the SMR as a multiplier and putting a weight of one on it is itself an arbitrary choice.

Member Needleman: And just to re-emphasize that, you know, you say, Susannah, these are small adjustments. Yeah, there's some judgment involved, but you made judgments. You've made a judgment that the weight for mortality was going to be one and the weight for nursing homes is going to be .5. And I can ask the question, why isn't the weight for nursing homes one and why isn't the weight for mortality .5?

And there's no empirical analysis of the consequences of those weights or how they relate to direct measures of other stuff. So again, I go back to its -- Patrick's argument is the impact seemed excessive. My argument is, whether it's excessive or not, the decision to weight them at one and .5 are completely arbitrary.

And I'd like to see a lot more analysis of the both the impact and how sensitive the measure is to alternative weighting strategies, and some accurate, more detailed discussion of why the weights were chosen that were chosen. And I'm not seeing that.

Dr. Herrin: So, yes, I will agree that, as you're saying, that one is a choice for a weight. I think that's -- well, I'll back up a little bit.

First of all, you know, ideally, we would somehow model all of these things at once so that there was no choices to make at all. And we tried a lot of ways to do that, sort of different time varying models and censoring, and all those things, there were lots of problems with them in structuring to count days that occur.

So, yes, we made a decision to calculate the days using an offset and then to adjust them. Well, why use one? We make decisions all the time when we do statistics about, you know, weighting things, and we often use the weight of one when we're averaging things, because we think things are the same.

We did look at a range of weights. We looked at the impact on using different weight choices. When you see an extreme value of, you know, 10, you mentioned in the table correction, well, that's not because the SMR is very high, but because that the actual expected days in care was very high. So, what that means is that ACO was already an outlier. So, yes, so it's more of an outlier.

You know, when we looked at changing the weights, we saw that scores shifted, but there was very little change in the ordering of the scores. And I think that's what gave us some confidence to move ahead with the weights we have.

I think that, you know, as Susannah said, our primary goal was to provide some kind of, you know, performance assurance that we weren't

ignoring mortality, especially when incurred at very high outlier rates at ACOs.

But I think that, you know, there's not a -- there's no standard to use to decide what the right weights would be. And I think, you know, you have to make some decisions, and, you know, we did look at a lot of range of weights and combinations and see if they didn't shift. And, you know, we did bring this problem, this specific problem, to a discussion with our TEP and we felt that, yes, you know, it made sense, mostly sort of from a more of formality perspective, that nursing home risk would count less than a mortality risk because it's a less severe outcome to patients, and, you know, half was a reasonable thing to use.

But, yes, certainly on days, these are -- you know, I won't say they're arbitrary. I think one is a pretty typically chosen weight, but they are subjective, definitely.

Dr. Bernheim: And using expert input, right?

Dr. Herrin: Yes.

Dr. Bernheim: And we've looked at what the difference is if we change them, and it's not a big difference.

Dr. Herrin: You're on mute, Patrick.

Member Romano: I'm sorry. I think Dr. Perloff has a very important comment. But, just out of curiosity, did you -- is it possible -- because she has another idea of about what you could have done. I mean, I guess there are ten different ideas about what you could have done instead of this approach you did take, but could you just adjust the offset term in the negative binomial?

In other words, if the SMR was two at an ACO, and so that translates into that the average patient should have had, you know, two extra days of life,

could you just add those two days to the offset term in the negative binomial? And, therefore, directly account for these excess dead days, so to speak.

Dr. Herrin: I mean, if you think that -- am I muted? No. I mean, you could do that, but it's not -- you know, it's sort of an arbitrary choice, then, right? I mean, an SMR is not counting days. It's a ratio.

I would like to point out that the SMR we use is actually -- we calculate patient level SMR, basically, thee particular expected mortality. So, it's only going to be high for a patient who has a much -- you know, that'll tend to be similar across the ACOs because -- I mean across a given ACO. But, yeah, what you're suggesting, you know, would be, you know, an SMR is at a ratio trying to expand, use it to adjust the offset and require some kind of --

Member Romano: Yeah, it just seems a little awkward to apply a ratio to a difference measure. But, you know. Dr. Perloff.

Ms. Elliott: Yes. Jen, did you want to share your comment that you put in chat? She may be having audio issues, so I'll read her question.

"I really like the mortality model approach, but when we do this in pricing work we use the probability of death as a covariate in the PMPY RA model. Is this a reason this type of stage modeling was not possible? As a side note, I'm not sure transitioning to long-stay nursing home status is always negative, but could be convinced otherwise."

Dr. Bernheim: You can respond to the first part, Jeph. I'll respond to the second part.

Dr. Herrin: Well, yeah, I mean, we're trying to capture something about the excess risk of mortality due to being enrolled in that ACO. So, that's fine. We did not try to, you know, put in a patient level probability. I guess you can because, you know, a ACO may have -- yeah, so, you'd have

to disentangle those two features: the patient level excess risk due to the risk factors and the ACO extra risk and trying to account them separately. But then -- yeah, I don't know how that would work out in practice.

Dr. Bernheim: Just about nursing home status being always negative, we agree, right? So this was part of the challenge of constructing this measure, right? What we didn't want to do is construct a measure that meant that anytime a patient was transitioned, that was necessarily a negative. And the TEP was really interesting on this topic. We spend a lot of time on the topic because, you know, there's a strong goal to keep people at home as much as possible. But it's also true that that is not best for every patient, depending on lots of factors.

And so we didn't want this measure for an individual patient to discourage that transition if it's the right thing. And we didn't want to encourage it. And so, that's how we landed here, was to have -- you know, again, so the goal is, if you're already in long-term nursing home care at the start of the performance period, there's essentially no -- there's no penalty.

Penalty is not the right word, but I'm blanking. And if you transition, it doesn't count against your days as days in care. But if you've an given ACO has much higher rates of transitioning than would be expected given the case mix, there is this little bit of a weighting.

So it's trying to balance the sense that we don't want this to be a measure that encourages -- the counter is really a lot of work that CMS is doing, quite honestly, right now, to try to build systems and support to keep people at home, but we also didn't want to over-go the other direction.

So, is it perfect? No. Was it done with a lot of thought with policymakers and experts? Yes.

Ms. Elliott: So at this point, the discussion has run a little bit long. So Dave and Christie, if I could loop you in to present next steps on this measure?

Co-Chair Nerenz: Yes, well, I guess there's still -- we could leave the window open just a little bit for any other members of the sub-group or the SMP in general to weigh in.

Although, I think the fork in the road we get is is there going to be a re-vote or not? I have to look for guidance here. I think on those who have been active so far, Patrick, Jack, Jennifer. You've heard responses, you've raised questions. The --

What would you like us to do? And maybe there could be a chat response, or you could just mention it. But I guess for -- we should allow anyone else who hasn't spoken to do so if he or she wishes to do that. And then, we hit the fork in the road. Okay, silence is silence. That's fine.

Ms. Elliott: There is a chat, David. Sorry, this is Tricia. There is a chat from Jack. He'd like to re-vote on validity. And Sean O'Brien is saying from sub-group number 1, member bottom line, it's a composite measure should be assessed, interpreted as a composite. Discussion did not change my assessment of the validity, according to Sean O'Brien.

Co-Chair Nerenz: So we're all over the place. The composite issue strikes me as tremendously difficult, because it changes everything. And I have really no idea how to go down that road. I do think it's straightforward to re-vote validity.

And I think actually for highest level of closure, it probably makes sense to do that. So even if people feel that their opinions have not changed, they can indicate that, and on we go.

So I guess my inclination is to say yes for thoroughness and completeness. Let's revoke

validity. If it changes, it changes. If it doesn't change, it doesn't change. The composite discussion is a deep, big discussion.

Member Romano: Well, it's tied together, right? Because if certain people, and I made the argument that because these weights are arbitrary, that it effectively is a composite. And that it hasn't been presented as a composite.

So if the goal were simply to do a mathematical correction of the days in care, to account for excess deaths and excess dead days, if you will, at certain ACOs.

That's a mathematical correction that's based on a mathematical relationship and formula. This as we've been discussing it is actually a composite measure.

And so it does raise the question of how can we support the validity of this measure as it's been presented to us when it hasn't been presented as a composite?

Co-Chair Nerenz: Yes. No, and Patrick, you were very clear on that. And I follow that exactly. My pointing out of the complexity is just the process complexity for us, that we just have a different process and rating criteria for a composite measure.

There would be no alternative but to table it. And, you know, I don't know, do some kind of offline re-vote or something. If the consensus of the group is it is in fact a composite measure. I just, simply re-voting one dimension won't do the job, because our criteria are different for composites.

Dr. Bernheim: I think, can I just ask a clarifying question, and Tricia, this may be to you. So I think of a composite measure as taking multiple standalone measures that are independent quality measures.

And per our discussion for instance, I don't think people want the transition to nursing home to be judged as a standalone measure.

It's designed as a balance, a balancing weight in this measure. It's designed to be there to guard against unintended consequences in extreme cases.

But it's not; it's not designed as a standalone relative performance measure of the ACO. So we may put you in a little bit of an awkward position.

But so we really didn't build this as a composite measure. It couldn't, it couldn't stand as a composite measure based on my understanding.

We're not trying to combine three pieces of information about quality; we're trying to balance an important piece of information about quality with some related information to prevent unintended consequences of measurement.

Member O'Brien: This is Sean. Sorry, I didn't mean to throw everyone off their rail. I wasn't suggesting to reclassify this as a composite or continuing discussing it as a composite.

I was explaining my concerns with the measure as presented. So I submit concerns about the validity and interpretation, and I thought, I didn't mean to bring up a new point.

I thought I was just kind of echoing, I don't know, my wording wasn't perfect. So I would say just simplest to disregard my comments and keep moving forward.

Co-Chair Nerenz: Yes, and if we're, for what it's worth, my sense of the core essential definition composite matches, what Susannah just said, it is the combination of what otherwise could or are, could be or are standalone quality measures.

And that either is or is not the case here. Susannah's point obviously, it is not the case. But

what's a composite measure? That's my understanding as well.

Dr. Herrin: And composite measures would require us to choose arbitrary weights to combine them also, right?

Member O'Brien: Well, part of my point though, and it's also been made by Jack and Patrick and others is there, is that there's a weighting that's going on. And so, and it is arbitrary.

And you know, the major description sounds like it's kind of a risk-adjusted measure of average days at home. But under the hood, it's actually, there's reasons why straightforward measures of days at home, literally it maybe is not desired, but because it's not desired, you want to incorporate adjustments. The end result is something more complicated, that's some kind of combination of three things rather than an underlying single thing.

Dr. Bernheim: It feels sort of like if we had buried it in a single model, the committee would have no problem with it, right? I mean we adjust for all kinds of things in our measure. We've made this explicit.

But you know the results are really driven by the days at home. I mean you can look at what the impact is for 90 percent of the things. I mean it is the days at home, like the days in care, that are driving the ACO results. That's the primary measure.

I mean again, I, I appreciate the complexity this brings. One of our TEP members said they thought the complexity was great, because they thought it would make it harder to game.

So the ACOs would know that, you know, couldn't game anything, mortality or nursing home, but they wouldn't figure out how to game it.

But there are lots of perspectives on when complexity is useful. But again, it is, it's not so different than what we do with lots of our measures. It's just made more explicit.

Co-Chair Nerenz: And we, a meeting or two ago, we did have a very detailed discussion about this gray area of what's composite and what's not. And if you put two or three things together in some mathematical formula, does that inherently make it a composite?

And I think we said no. It's really got to be this combination of what are or could be standalone measures. And I think what I'm hearing clearly is from the Yale Core perspective, this is not being presented to us as a composite measure.

They're not intending it to be a composite measure. So I guess we would have to take the view that no, in fact that's not correct. That's a pretty high bar, I think, for us to reach.

Let me make a proposal here just to move this along. There has been a suggestion that we re-vote validity. I think we can do that. The question of whether it is a composite of, there's maybe there's a staff expert adjudicator who could filter through this.

I think clearly where we are is that the developer here presents this to us not as a composite measure. And they've made an argument that in fact it is not a composite measure.

I don't think we have time this afternoon to get into any kind of discussion with it as if it were. In fact, we'd have to go start the review process all over again. Just can't do that this afternoon.

So I think the practical discussion would be go ahead, re-vote validity on the presentation that's been made to us on the characterization of the measure that's been made to us. And if there's any

lingering concern about composite, I think we just have to pick it up after the close of this meeting. I just don't know how to work it in in the time we have right now.

Member Kaplan: David, this is Sherrie, can I just ask for the staff, is the requirement for a new measure, it's face validity only? Is that accurate? I mean that's the, you have to at least demonstrate face validity? Beyond that --

Co-Chair Nerenz: Right.

Member Kaplan: You have to at least do that. But you don't have to do what the Yale team in addition has done?

Dr. Pickering: This is Matt. That's correct. New measures, the minimum validity assessment is face validity. Developers can also do empirical validity testing at the different levels, patient level encountered, the kind of entity level. But minimum acceptable is face validity for the new measures.

And relating to the composite discussion, David, I agree with your approach. I think if, you know, if this was to be deemed as a composite, it would have to be resubmitted as a composite measure and evaluated against our criteria as a composite measure.

Which I believe that Yale has articulated, and this group has made an agreement to not go to the composite realm. But just for clarification there, that it would be able to vote on this measure as is as a composite, because it's not submitted as a composite.

So just noting that. So I think if theirs is a interest in re-voting on validity, maybe kind of articulating why that re-vote would take place. And then, moving to the vote.

Member Romano: It's already been clearly

articulated, and Jack confirmed in chat that it's an issue of the face validity of the construction of the measure. And the link of the risk adjustment models, which is also part of the domain of validity.

Dr. Bernheim: And was part of what the TEP was supporting in their face validity vote.

Dr. Pickering: I'll just check with the team to see if we'd like to move to a vote based on this. And I'll just call out the comment from Alex. Thanks, Alex, for mentioning these items.

We should consider all the information presented even if it goes beyond the minimum required. That's correct. Based on the conversations we had yesterday, you know, there's those different types of testing that can be considered empirical testing approaches.

But if there's other testing that's been provided, it should be considered, and that may change your overall rating based on all the information that's been provided to you.

Ms. Elliott: Matt, we do have one hand raised, Larry Glance.

Member Glance: Thanks. Just a quick comment. And I think, Matt, I think certainly your understanding and interpretation of the algorithm is extremely helpful. At the same time, I think we have to sort of put things a little bit in context of what our usual practice has been as a panel.

And my understanding has been that even for a measure that is a first time measure, certainly what we've done is we will substitute face validity for empirical validity testing. But the other threats to validity are still relevant.

So for example, if the risk adjustment model is not valid, then that is considered -- in my experience, it has been considered enough of a threat to validity

to invalidate the, to make a decision that the measure is not valid.

So it's not just about looking at the face validity. We also look at threats to validity. And if there is a threat to validity, like a poor risk adjustment model, then that will shift the decision over to making the measure not valid. Thanks.

Dr. Bernheim: And are there concerns about the primary risk adjustment model for the excess days in care? We didn't hear any concerns about the risk adjustment model.

Member Glance: Well, I thought we just did. We were talking about the fact, the way that the risk adjustment models were constructed and put together.

Dr. Bernheim: I think those really adjust to what we sometimes refer to as the additional adjustments to the overall measure based on the competing unintended consequences of mortality in nursing homes. But the primary model is the excess-days-in-care model.

Ms. Elliott: Okay, David, can we turn it back to you for next step?

Re-vote Measure 3667

Co-Chair Nerenz: I don't think my name is Solomon either first name or last name. But I think we ought to tee up the re-vote on validity. That seems to be the action step that has been requested. And we can do.

Ms. Elliott: Okay. Hannah, do you have the vote ready?

Ms. Ingber: I've sent out the voting link to all those in sub-group 1. So those who are in sub-group 1, if you could pull that up. I will open the poll. So voting is now open on measure 3667 for validity, your options are high, moderate, low, and insufficient.

Ms. Elliott: I think we're waiting for two more votes to get to quorum. I think we're there, I'll just double check the responses. Quorum is nine people though, yes.

Okay. I'll now lock the poll and share the results. So for voting is now closed for measure 3667 on validity. We have zero votes for high, four votes for moderate, five votes for low, and one vote for insufficient.

Therefore -- just double checking my numbers. Member Needleman: Should there have been 11?

Ms. Ingber: Quorum is ten folks. Therefore, this measure does not pass on validity.

Co-Chair Nerenz: Is it CNR or no-pass?

Ms. Ingber: I'm so sorry. You're right. Consensus is not reached on validity for 3667.

Co-Chair Nerenz: And it is an important distinction.

Ms. Ingber: Yes, my apologies. Okay, I can stop sharing the results now. Thank you, everyone.

Measure 0689

Ms. Elliott: Thank you. The next item on the agenda is we're going to continue a discussion on 0689. David, do you have any, or Christie, do you have any opening comments? Or we can, we have NQF staff, and they're ready if you'd like a refresh of the overview of the measure.

Co-Chair Teigland: Yes, we can do a refresh. And then, I may have some comments. I've reviewed all of the submission materials again, myself. I'm sure others did too. Thanks.

Ms. Elliott: Okay. Poonam, I'm handing things over to you for a quick refresh of the measure.

Ms. Bal: Can you hear me okay?

Ms. Elliott: Yes, we can.

Ms. Bal: All right. So just wanted to remind everyone of our discussion yesterday for 0689, percent of residents who lose too much weight. We wanted to make sure that we did have the right submission data in front of you.

We've heard that there was some confusion with the way that we sent documents this cycle, since we have switched over to our new database. We are attaching the old testing attachment that is now merged into our major database.

So we just wanted to make sure the additional data was provided to you from 2015 to, just for your information. We really want the data that was from 2018 to 2019 being used per your review.

Just a reminder that for this measure, it did pass on reliability. And then, based on the re-vote yesterday, it did not pass on validity. This measure is a maintenance measure.

So it has been previously endorsed and is an outcome measure. It's assessed at the display level, and it was not risk-adjusted. The data source for this measure is the minimal data set, MDS 3.0. And the collection instrument is the resident assessment instrument, RAI.

The developer did conduct two forms of testing for this measure for validity. There was critical data element testing, which relied on previous studies that looked at inter-rater agreement.

There was no concern expressed by the SMP about that level of validity testing. The concerns really came in at the performance score level, which was using a correlation with other measures of Nursing Facility Quality, including facility CMS, 5-star rating, health inspections, rafting, and staffing levels, for overall and RNs.

Validity testing was conducted. The correlation results show negative correlations between the facility level weight loss, QM score, and the overall quality health inspection rating in RN staffing.

There were some concerns that the results had weak negative correlations between the facility level weight loss QM score and the overall quality rating.

And then, one where you didn't know that the low correlations are common, especially lower when typically seen to indicate an overall nursing home quality and staffing may have very little impact on residents' likelihood of losing weight.

Seasonal variation was also tested. The seasonal variation showed highest weight loss in Quarter 1 with progressively lower rates in Q2 and Q4. There was also concerns raised about the developers did not risk adjust.

But the developer confirmed that they did explore risk adjustment, and that, but their attempts to develop a risk adjustment model was unsuccessful mainly due to low R-squared values. So that's just a quick reminder of the measure and what we're looking at.

Co-Chair Teigland: So I can start, and I'll just comment that the reliability statistics did change dramatically from the previous testing document. It does look solid at .76.

Validity also looks a little bit better. But I still have some concerns. It, definitely the measure, you know, it has, seems to have good, good stability across you know, across both -- he presented the data multiple ways across states, across facilities, etc.

The correlations with the related quality measures are a little bit stronger. They're in the expected direction. But I don't necessarily agree that demonstrates strong validity. They're really still

very, very small, very weak correlations, which is, I do get that that's, you know, common to see.

We do see definitely meaningful differences in scores now across facilities, much more, you know, divergent scores than we were seeing in the earlier testing data. And we don't see those big shifts.

I see that now from, you know, a large portion of facilities moving three deciles or more from quarter to quarter. My biggest concern still remains that you stated in your document that you didn't -- no ICD-10 diagnoses were considered for risk adjustment, were even considered for risk adjustment.

And you say that you base your, you know, your review, your models on a review of the literature. You know, I think a review of literature, of the weight loss literature, suggests that there are many chronic conditions that can, are associated with weight loss.

Like depression, like Alzheimer's disease, cognitive impairment, Parkinson's disease, you know, cardiac disorders, eating dependencies. There's an item on the MDS called leaves 25 percent or more of food uneaten. You know, swallowing or chewing problems. I just, I'm confused as to why you didn't --

You just said we're not even going to test those at all. Also, well, you did find that age, you tested age and race. Race was not significant, and that makes sense to me. Even though there's some literature showing that there is a disparity in weight loss. But that's not showing up in this data.

But there was a pretty strong impact of age over 85 in at least 17 percent more likely to lose weight. You say that your C-statistic was only .5 something, and that was just too low.

We see less of C-models with C-statistics at .5 or less that they're not, they're frequently not very,

very high. So yes, again, why you chose not to at least adjust for age?

There could certainly be facilities that have lots and lots of people who are much older and could be impacted by that. But I'm more concerned about the fact that you didn't even test any of the chronic conditions at all in the risk adjustment model.

So I have concerns about the model. Anyone else on sub-group 2 have any additional comments on this? So I am not so sure my vote would change on validity. I don't know what the other subcommittee, you know, members think after reviewing.

Hopefully, we all now have all of the right data in front of us. Do we need to re-vote this? Or did any of the new information -- if you weren't looking at the wrong information, I was, I admittedly was -- Change your mind? And I don't see a reason to re-vote this. But if others do, I think we could take time to do it.

Dr. Nagavarapu: And Dr. Teigland, this is Sri from Acumen. We have results that speak to your questions about risk adjustment, where we have tested Alzheimer's, dementia, depression from MDS items.

Those do have odds ratios that are slightly above one, but they're small. And the C-statistics from those predictive models are extremely close to .5. And I should note that for a C-statistic, you know, .5 is a coin toss. So this is not the same thing as an R-squared.

Co-Chair Teigland: And --

Dr. Nagavarapu: And so .51 is essentially a coin toss for a positive predictor. So for Alzheimer's, dementia, and depression, we conducted this testing. You do see a positive effect, as you would expect.

As I mentioned, it's small. And if you look at the impact on measure scores, about 97 percent of facilities stay within the same decile ranking. We also tested some of the other covariates that you mentioned when we saw this comment from the SMP.

The reason for not testing these prior is that there are very strong conceptual reasons for not including them, which I can get to in just a moment.

But the covariates that we included were mechanically altered diet and no-swallowing disorder to get at some of the ideas that you're talking about and also prove further the validity of the data element.

As I mentioned yesterday, for mechanically altered diet, the odds ratio is, as you'd expect, greater than one, 1.61, and significant. For no-swallowing disorder, the odds ratio is, as you would expect, less than one, .47, and statistically significant.

In both of those cases, the conceptual reason for not including these as risk adjusters in the measure is that these items are items that CMS has observed a responsiveness to in terms of the payment incentives inherent in the patient-driven payment model.

And so risk-adjusting for these items in a case where CMS is beginning to surge in coding is potentially problematic and can lead to inappropriate incentives for coding that exacerbate the incentives that we're seeing now.

And even if you adjust for these though, if you construct the measure using these covariates, you still see that about 20 percent, about 80% of facilities remain within the same decile.

So this is all to say that the risk adjusters that you have in mind do work in the way that you would expect to justify that this is a valid measure and to

reinforce the validity of the data elements.

For the ones that do not have conceptual issues with them, they have negligible impact on the provider rankings. For the other ones, that have a slightly larger impact, there are serious conceptual reasons that CMS prefers not to adjust for them.

Co-Chair Teigland: Yes, so what your submission form says is no ICD-10 diagnosis from item 18Y8000 were considered for risk adjustment. So that's, I think, what we have to go on what was in the submission form.

I don't have that information in front of me. No one on this committee does, which is unfortunate. Yes, I mean, C-statistic is debatable. CMS has models that, they're in the 5-star rating system, that have a C-statistic of .3. So we won't go there.

But yes, I don't know what, if any of the committee members have any thoughts, or if not on our committee, open it up to the whole SMP.

Dr. Nagavarapu: I should note that a C-statistic ranges from .5 to one for a positive predictor.

Co-Chair Teigland: Yes, I'm thinking of -- one, readmission. I'm not seeing any --

Ms. Elliott: Larry, I think Larry Glance has a hand raised, Christie.

Member Glance: Sure, thanks.

Co-Chair Teigland: Yes?

Member Glance: Could the developer just summarize very, very briefly, other than the comorbidities or process variables that were felt to not be valid for inclusion in a model, what comorbidities were actually examined in a final model? Did they create --

In other words, did they try to create a non-

parsimonious model that would include a large number of comorbidities that might be associated with unintended weight loss? And what was the performance of that model?

Dr. Nagavarapu: Yes, thanks for the opportunity to talk through the results. This is something where when we submitted our developer response, there was uncertainty about whether the SMP was reviewing the right forms. And so we constructed these responses after the developer response, which is when citing them verbally.

But to answer your question, we estimated a model with mechanically altered diet, no-swallowing disorder, age being greater than 85, and three comorbidities that were raised both in the literature as well as in some of the SMP comments that we saw.

Those were the presence of Alzheimer's, dementia, and depression. So that, that's the model we tested. We tested one model with all of the covariates I just mentioned. And then, one model removing the ones that were more problematic and retaining only Alzheimer's, dementia and depression.

The model with all the covariates I mentioned has a C-statistic of 0.6. That one is the one with mechanically altered diet and no-swallowing disorder, where CMS has observed these increases in coding associated with the new payment system.

The model without the, model with just Alzheimer's, dementia, and depression, it has much less predictive power, and the C-statistic drops to 0.5.

That's the case where 97 percent of facilities have a ranking that stays within the same decile if you construct a risk-adjusted measure based on those three comorbidities versus none.

Member Glance: Thanks very much for that. I guess, based on your presentation, I'm not really

convinced that you've explored all the potential associations between comorbidities and the outcome of interest.

And I think, you know, that the model that, the final model that you came up with, who has two or three covariates, understandably has a low C-statistic.

And I think it would probably make sense to explore a more non-parsimonious model first before ruling out the possibility that risk adjustment will make a difference in terms of how nursing homes are classified.

That was the main issue that I think a lot of us had was the lack of risk adjustment. And I'm still not really convinced that you've made a strong argument that this outcome should not be risk-adjusted.

Dr. Nagavarapu: So our testing for risk adjustment really focused on sort of the conceptual model for the key factors that can influence weight loss and at the same time would be factors that the facility may have a more difficult time compensating for.

And so that led us to focusing on these specific covariates. I think the worry was that if you throw a large number of risk adjusters that may not be as clinically meaningful --

And I think the ones we picked were clinically meaningful in the sense that those are the ones that SMP really focused on in their comments as well.

But if you sort of throw a lot of things in that are not necessarily clinically tied, you get the threat of sort of data mining in a way that gets away from political appropriateness of the risk adjusters. And so that was our concern.

I should note specifically the case of the cancer item. This was brought up in the SMP discussion, I think. The cancer item actually is one we are

interested in testing.

The reason we did not is that the specific MDS item that would be best to use there is one that was added late in 2019. And so testing with the cancer item would require using 2020 data. And there's a concern about using 2020 data in the testing due to the pandemic.

So the cancer item is a thought from SMP that we noted, and we agree with, is interesting, and important to test. And we anticipate doing that in the future. Because of data limitations, it's difficult to do that now.

Ms. Elliott: So Christie, this is Tricia. There's a couple comments, including one of yours that came in through the chat. Jack, do you want to share your comment?

Member Needleman: Sure. Now, I was just reacting to the reaction about the risk adjustment model not showing a strong, not weighting up Alzheimer's, depression, and dementia.

And my thought was that nursing homes that provide good care may well have met the challenge of assuring that they can effectively meet the nutritional needs of those patients. And if that has become the routine standard of care, it would not show up strongly in a risk adjustment model.

The question is what. And this turns on the cancer discussion we were just having. Diseases that might be associated inherently with a wasting of the patient, with weight loss, probably should be in the risk adjustment model.

And it's not clear how, you know, what? I'm not a clinician. So I don't know which conditions meet that characteristic and whether patients with those conditions stay in nursing homes once wasting, once weight loss starts for them because of those conditions.

And those who have clinical issues that perhaps the steering committee on this one, with more clinical expertise, could do a better job of assessing.

Ms. Elliott: And Christie, did you want to share your comments that you put in the chat?

Co-Chair Teigland: Yes. I was thinking R-squared, of canceling the models that have R-squared. Of course, the statistic is .5 to 1, I misspoke. Yes, and I can think of an example, Jack, like Parkinson's disease.

You know, even if you're managing it as well as you can, people end up taking multiple, multiple pills a day, and that interacts with food.

And it makes it difficult to plan your food or eat food. And there's not a whole lot, if you're going to manage the symptoms of Parkinson's disease, that facilities can do probably.

People might still lose weight. And I don't, yes, maybe that, you know, I agree that some of these conditions, maybe you don't want to adjust away, you should be dealing with depression.

But I don't know if you can prevent all weight loss in people who have advanced Alzheimer's disease, for example. Yes, that's a --

Member Needleman: I don't know either. But these are --

Co-Chair Teigland: Yes, yes.

Member Needleman: You know, it's thinking through the clinical experience of the patients. And what's possible --

Co-Chair Teigland: Right.

Member Needleman: -- that ought to govern what goes into the risk adjustment model. We've been seeing with large data sets some, you know, kitchen

sink models, standardized CMS models.

Co-Chair Teigland: Right.

Member Needleman: With lots and lots of variables and we prune them down only when we absolutely have to. But here we're asking the developers to think through in a more careful, thoughtful way.

Who's in the nursing homes and the clinical pathways by which they might lose weight and whether those are preventable. Because it's the unpreventable weight loss that we want to capture in the risk adjustment.

Co-Chair Teigland: Right.

Member Needleman: Okay.

Dr. Nagavarapu: And yes, we totally agree, Dr. Needleman, with sort of the way you're thinking about both the Alzheimer's dementia and depression effects that we're seeing. Where we're seeing these very small effects.

So even though they go in the direction you'd expect, they tend to be quantitatively small. And that very likely has to do like, like you're saying, with a compensatory response by facilities that we would want to see.

I think we've been able to address each of the covariates that people have. All the salient ones that people have brought up on the discussion and in the points. The one that we didn't talk about is one that Dr. Teigland mentioned about the 25 percent or less in terms of nutrition.

We looked into this MDS item and just want to note that that particular item is only filled out for a subset of residents who are on parenteral IV feeding, or a feeding tube. And that 25 percent or less is an option for them.

So it's not an item that we have available across all

residents. It's an item that is associated with a very specific set of service provision decisions that a nursing home is making.

And so that makes it difficult to test that item and potentially misleading. But it's something that we could definitely keep in mind if the MDS items change in the future.

Ms. Elliott: So, Christie -- oh, I'm sorry.

Co-Chair Teigland: Oh, no. I was just going to say, it's been a while since I worked with the MDS. I can't remember if there's some eating dependency items.

Yes, I just can't remember what all the items are. It seems to me there were quite a few things related to weight loss. But sounds like you've looked at all the potential candidates.

Ms. Elliott: So I think we have one more hand raised. And then, maybe we can go back to you, Christie, after that, just to move forward if we need to discuss re-vote. So Patrick, you had your hand raised?

Dr. Pickering: Yes, I mean, I wasn't in this sub-group. So my question is really a process question. Which is, I think the risk adjustment issues here were extensively discussed yesterday.

I think that if I understood correctly, the issue at hand was that some people might have been looking at the wrong information with respect to the correlation testing results, that is the measure score of validity testing results.

And so, I guess, if there were any minds to be changed, it would be around that question of whether the newer data related to a measure score validity based on correlation testing.

Is, it gives a more persuasive results than whatever the previous data were that people were looking at.

I'm just trying to understand the process to make sure that this is being reconsidered based on some new information or correction of the information.

Co-Chair Teigland: Yes, and a little help from the committee. I don't know if other people were looking at the previous testing form. The results on the correlations were pretty similar. So I, you know, I don't, I didn't see much, much change there.

Yes, the risk adjustment model, I mean it still says in their testing form they didn't test any ICDs. So we don't have that information in front of us. I guess, and did you provide that, all of that data? I may have missed it. In the response that you just described to us?

Dr. Nagavarapu: Yes, for the risk adjustment results, those were items that came in after the developer response. For the developer response, we were really trying to focus on specific things to clarify.

Co-Chair Teigland: Yes.

Dr. Nagavarapu: Like any confusion with the forums and any very basic points that were associated with that. The one thing I would mention in terms of the correlations and the convergent validity.

Which also the summary of the measure today did not mention is that one of the key concerns noted in the SMP discussion was that the correlations were done only with star ratings, things like the staffing rating.

And that's actually not true of the forms that we submitted. In the forms that we submitted, the correlations are also done with other quality measures, and the correlations are higher than the those with the star ratings as you would expect.

Now, those correlations all go in the right direction. They are not, you know, correlations that are, that

are very high. Right? They're correlations on the order of .15. But they're meaningful.

And they're not high by design in the sense that the nursing home compare program is curated such that the quality measures are not duplicative of one another. And so in cases where there would be extremely high correlations, likely there would be discussion of removing measures.

So that, our forums do have that additional information on convergent validity that was not discussed yesterday, or it was discussed inaccurately yesterday.

Re-vote Measure 0689

Ms. Elliott: And Christie, there's no more hands raised or chat. And we are bumping up against time for public comment. So we have to ask for your decision on whether or not to re-vote.

Co-Chair Teigland: Yes, yes. And with lack of any further guidance from the rest of the committee, I think, just for my own personal gratification, we should do a re-vote, quick.

Because we did get some new information yesterday that at least one of us or some of us or more of us had not previously looked at. So can we just do that quick?

Ms. Elliott: Sure. And we do have support of that re-vote from Alex Sox-Harris in the chat.

Co-Chair Teigland: I just see it, yes.

Ms. Elliott: Okay. So we'll go ahead with a re-vote. Hannah, can you initiate the vote? And quorum for this measure sub-group is eight. Hannah?

Ms. Ingber: Yes, thank you. Voting is now open and on measure 0689 for validity. Your options are high, moderate, low, or insufficient. And we're using the same voting link that we sent out for sub-group 1.

But this is just a sub-group 2 measure.

Okay. We have nine votes, which is what we were expecting. I will share the results. So voting is now closed on measures 0689 for validity.

We have zero votes for high, four votes for moderate, five votes for low, and zero votes for insufficient for a total of nine votes. Therefore, consensus is not reached on validity. Thank you, everyone.

Ms. Elliott: Great, thank you, Hannah. Okay. So next on the agenda, I'm going to look to -- while I'm catching up on the agenda here.

Given the time is 4:45, Dave and Christie, I would recommend that we move to open up for public comment. And I think we're going to have to defer some of these SMP topics to a future meeting. Would you concur?

Member Needleman: Yes, I would. And we've got some really brief updates, but those could be done by email. The other things cannot possibly be done in five minutes.

Opportunity for Public Comment

Ms. Elliott: Okay. So we'll move to public comments. And then, we'll wrap things up. Okay. So at this time, I am opening the meeting for any public comments. You can enter those into the chat.

Raise your hand if you're on the line, the WebEx for the hand raise feature, or unmute your phone line to offer any public comments. At this time, I do not see any hands raised or chat. I do not hear or see any open phone lines. Give it maybe one more minute.

Okay, I think we have allotted enough time to give folks the opportunity to unmute or chat. Okay. Can we go the next slide, please? Okay.

Next Steps

So some next-step reminders related to the measures reviewed today. The -- oh, actually, Hannah, I think I'm stepping on your lines. Why don't you go ahead? Sorry about that.

Ms. Ingber: It's okay, no problem. Yes, so some reminders about next steps for this cycle, the Fall 2021 cycle. So full measure submission deadlines are coming up on November 1st, 8th, and 15th, depending on what group the topic area is in.

And for the SMP's discussions yesterday and today, NQF staff will summarize the relevant measure information into a meeting summary and provide that to the various standing committees.

So they'll see this, a summary of the discussions from yesterday and today before, while looking at all of the other measure documents. So the standing committees will evaluate measures in a February timeframe.

And then, the Consensus Standards Approval Committee will review the Spring 2021 measures, the ones that we reviewed at the last measure evaluation meeting, on November 30th and December 1st. And for these measures that we discussed, they will be reviewed by the CSAC in a June or July timeframe.

Also for the Spring 2022 cycle, the next intent-to-submit deadline is January 5th, 2022. So we encourage all measure developers who are looking towards that goal to reach out to us if they have any questions or would like any technical assistance.

Next slide, please. So for our next SMP advisory meeting, that will take place on Tuesday, December 14th from 12:00 to 2:00 p.m. Eastern Time.

And we will discuss the CSAC updates that we have

and continue discussion of the reliability threshold table and continue discussions of policies for maintenance measures on requiring certain testing.

Next slide. Thanks. So as always, please feel free to reach out to us with any questions or comments at methodspanel@qualityforum.org.

You can also feel free to give us a phone call and the project webpage where all these materials are posted. And the meeting summary will be posted there too.

Is on, is the first link on this page in the SharePoint site where the panel members can receive, can review materials, is below that. I'll hand it back to you, Tricia.

Ms. Elliott: Thanks so much, Hannah. I'd like to give an opportunity to Dave or Christie for any closing comments as we get ready to adjourn.

Member Needleman: As always, thanks to everyone for all the thought and work that goes into doing this. Obviously, we can recognize and appreciate the time yesterday and today.

But what we don't always see is how many hours go into this behind the scenes. And careful reading and checking and thinking and commenting. And we just appreciate all that. It's a service to the profession, service to the field of quality measurement. It can't be done without all of your efforts. Really appreciate it.

Co-Chair Teigland: Yes, I second that. Great work on the SMP team side. And the developers, I'm just really impressed with, you know, your diligence, your hard work, your passion for your measures.

That came through loud and clear with all the developers we heard from over the last couple of days. So we appreciate that too. You know, we, we're all in, we're all in this together. We all want,

you know, good quality measures.

We think the ones that you've submitted are important. So for the handful that didn't make it through, we look forward to seeing them again based on all the great feedback from all the experts on this team. So thanks, everyone.

Ms. Elliott: Excellent. Sherrie Kaplan, I see you have your hand raised.

Member Kaplan: Yes, at the very, very grave risk of delaying departure here, I wanted to just ask did the developers have any opportunity to work on the recommendations of the questions raised during these reviews? And then iterate them back and forth? Or is it game-over and you got to resubmit?

Ms. Elliott: Poonam, would you be able to address that question more specifically?

Ms. Bal: Yes. So they don't have the ability to make updates to their submission following this SMP review. What we have, you know, submitted for reliability and validity has to be continued.

Just because when, during the submission, if they were to update that information, technically your review would no longer be valid.

But they do have opportunities to provide clarification, we do share a summary of this discussion with the standing committees, and there are other opportunities for them to share that information.

Member Kaplan: Thanks.

Adjourn

Ms. Elliott: Great question, thank you. Excellent. Everybody, eight minutes back in their day? Thank you so much. And I echo all of Dave and Christie in thanking everybody for their participation.

And also to the NQF staff that's done a phenomenal job behind the scenes to coordinate these meetings. Thank you, everyone. Have a great evening.

(Whereupon, the above-entitled matter went off the record at 4:52 p.m.)