NATIONAL QUALITY FORUM

+ + + + +

MEASURE DEVELOPER WORKSHOP 2016

+ + + + +

THURSDAY
MAY 5, 2016

+ + + + +

The Workshop met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 9:00 a.m., Wunmi Isijola, Moderator, presiding.

PRESENT:

KELLY ANDERSON, The Lewin Group
AMY BENNETT, JD, American Academy of Neurology*
SAMANTHA BERNS, MSPH, The Lewin Group
ALICIA BLAKEY, MS, American College of Radiology
KRISTEN BUTTERFIELD, MPH, Pharmacy Quality Alliance
KYLE CAMPBELL, PharmD, MS, Health Services Advisory Group, Inc.
PRIYA CHATTERJEE, MSPH, The Lewin Group
CINDY CULLEN, MBA, Mathematica Policy Research
LINDA DAILY, Livanta LLC
MARSIDA DOMI, MPH, American Healthcare Association
KAREN DORSEY, MD, PhD, Yale/YNHH Center for Outcomes Research and Evaluation (CORE)
TRICIA ELLIOTT, MBA, The Joint Commission
JEFFREY GEPPERT, JD, Battelle Memorial Institute
SHARON HIBAY, RN, DNP, Livanta LLC
WANDA JOHNSON, RN, MS, Oklahoma Foundation for Medical Quality
TONI KAYE, MPH, Physician Consortium for Performance Improvement
MEGAN KEENAN, MPH, Health Services Advisory Group
BRIGIT KYEI-BAFFOUR, MBA, The Lewin Group

LAUREN LEMIEUX, American College of Obstetricians and Gynecologists
KATHY LESH, PhD, Battelle Memorial Institute
LYNETTE LINES, MS, PMP, Yale Center for Outcomes Research and Evaluation (CORE)
COLLEEN McKIERNAN, MSPH, The Lewin Group
MAMATHA PANCHOLI, Agency for Healthcare Research and Quality
LYNN PEZZULLO, RPh, CPEHR, Pharmacy Quality Alliance (PQA)
SUZANNE POPE, MBA, American Urological Association
MATTHEW POPOVICH, PhD, American Society of Anesthesiologists
MELANIE SHAHRIARY, American Heart Association
SAM SIMON, PhD, Mathematica Policy Research
ZACH SMITH, MBA, American College of Radiology
KATHERINE SOBEL, National Committee for Quality Assurance
NAILA WAHID, The Lewin Group
ANN WATT, MBA, The Joint Commission

NQF STAFF:
HELEN BURSTIN, MD, MPH, Chief Scientific Officer
WUNMI ISIJOLA, MPH, Administrative Director
JASON JOHNSON, MA, Vice President, Information Technology
KAREN JOHNSON, Senior Director
ELISA MUNTHALI, MPH, Vice President, Quality Measurement
SARAH SAMPSEL, MPH, Senior Director
JEAN-LUC TILLY, Project Analyst
MARCIA WILSON, PhD, MBA, Senior Vice President
REVA WINKLER, MD, PhD, Senior Director

ALSO PRESENT:
TARA McMULLEN, PhD, MPH, Centers for Medicare and Medicaid Services*
MARY PRATT, MS, RN, Centers for Medicare and Medicaid Services*

*** present by teleconference**

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

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

9:05 a.m.

MS. ISIJOLA: Good morning, everyone. Happy Cinco de Mayo.

(Laughter.)

MS. ISIJOLA: Thank you again for joining us for day 2. We've covered some tremendous ground yesterday. We have a few items for today, but before we jump into that I just wanted to review again some of the items that we talked about yesterday.

Really yesterday was really informational, letting you know some of the work that we're doing here: measurement science, some of the things that we're doing and how your work impacts some of those areas.

Today it's really for the developers and how you interact with NQF, specifically around submissions. What we'll be doing in the morning is talking through some of our submission requirements. As we've heard in the past, there are some revisions based on your feedback that

we've implemented. And we continue to do that, but we wanted to showcase that to you again.

On our monthly webinars I know we give you bits and pieces of that, but I think this is an opportunity to really go through all of those upgrades, but also an opportunity for you guys to let us know if there is additional feedback or improvements that we could continue to make. So Reva Winkler and Sarah Sampsel will be talking about our requirements.

Our appeals and endorsement process. This was something that was approved by the CSAC, our Consensus Standards Approval Committee, and this is something we'll be rolling out shortly.

Also, I know I mentioned briefly yesterday about our off-cycle activities, another opportunity for our committee members to really continue to engage with us.

And then we'll go into I would call the critically acclaimed intended use project. This is something that will become more real in the near future. And Karen Johnson will be

talking about that, really some examples and how we implement that in our new criteria.

We'll also be talking about our CDP-MAP integration. So how does the endorsement process feed into our MAP selection process and how our efforts are being facilitated in the future in order for that to happen.

And then we'll have a summary of the day, but also an opportunity for you to continue to provide us feedback. During the last advisory panel one of the things we did was we got a group of you guys to really continue to give us feedback and suggestions on some of our policies and our processes. So that's another opportunity for you guys to engage with us as we continue to make improvements through all of our work.

So with that being said, I'm going to turn it over to Reva and she can kick off the day for us.

Reva?

DR. WINKLER: Thanks, Wunmi.

Who's running the slides?

Okay. We're not going to talk about absolutely everything, but we will want to talk about a few things that I think have been particularly of note in the last couple of months as we've started this whole group of projects for this year.

And so, the first one I want to talk about is something that is underused and perhaps because you're not aware of it. And so, we're trying to rectify that.

NQF staff will be more than happy to help you at any point. We call it technical assistance, but you can call it whatever you want to. It's pick up the phone and call us. We're willing and able to help you at any point in time with any questions you may have. And it doesn't matter as long as it's about measures. I don't know anything about football scores --

(Laughter.)

DR. WINKLER: -- but if it's about the measures, I'm happy to help. And so, if your new folks in your organization who are coming on

board -- we have new organizations coming on board, new developers coming on board. If you've got a new measure coming online, questions about what's happening with your maintenance measure. Anybody who asks, we get -- the phone rings all the time. And so, we talk to everybody. But it is underused by the measure development community.

And what we've found is that for those developers and measures that do come and talk to us up front early on, things go a lot more smoothly and we don't have sort of the train wreck of time crunched stuff happening at the last minute that drives us all nuts.

And so, I really want to invite you to take advantage of it. Even if you think it's a dumb simple question, I'd rather answer it well in advance rather than at the last minute when it may turn out to be bigger than you realize.

So what is technical assistance? And that's help in understanding the submission process and the requirements. And sometimes it's

a matter of what does this mean? Okay. We'll be happy to help you. Understanding the evaluation criteria that are used. For us, it's something we deal with all day and we know the long history of why they exist, but if you're relatively new to this or perhaps hadn't really thought about it, understanding why that's a criteria and why we ask the questions we do may help you answer it more on point. And so, we're happy to help and help you fill out the forms, whether it's technical or what words to put on the page. We're more than happy to help.

And we can do this at any time. If there's a project that your measure is involved in and it's already up and running, the Project Team is your first resource, but at any other time we are also available. And I'll give you a couple of examples of how this has worked out really, really well.

We had some measure developers come to us and say, gee, we're thinking about doing some measures and -- but they were uncertain about

things. And their first -- and they said would you take a look at our initial ideas? So Karen Johnson and I looked at their initial ideas and gave them feedback. And we then did that every six months for three years through the development of the process. We didn't tell them how to develop their measure. We just gave them a reaction thinking about our criteria, thinking about how we are used to seeing committees. When those submissions came in, they were great. They were great.

And it removed a lot of the headaches that I think we tend to run into because at the last minute the submission comes in, we're reading it and it's kind of like, ugh, this doesn't have the information we need that's going to help us get this to a committee. And so, we're back at you saying an I-need-it-by-tomorrow-at-2:00-kind of thing. That's all the stuff that drives all of us kind of nuts.

So if you have any questions up front, really, I strongly recommend that you come talk

to us. Sarah's over here nodding her head. Karen's over here nodding her head. It's just something that we see being so beneficial, but yet not -- widely underused. And we really want to encourage you to take advantage of it.

Now in terms of how do you get -- how do you contact us? All right. As I said, if you're in the middle of a specific project, that project team. And the contact information is on the project page. However, the default is always to go to your favorite page, Submitting Standards. And if you don't know this page and don't have the shortcut on your computer, think about it. You will see as you -- not too far down the page, Technical Assistance. All right? And it tells you -- gives you the email address. And we will triage it to the appropriate person in house and get back to you and help you out with whatever your issue is.

So, I mean, again, it really helps us, which is why we're really trying to emphasize how useful it is. And I think the developers who

have taken advantage of it have found it to be very, very useful. So for that; a plug, technical assistance is always, always available.

Any questions about that from anybody?

(No audible response.)

DR. WINKLER: Okay. Good. All right. The next thing we wanted to talk about is something that's ongoing, and hopefully all of you are aware of the fact that with this current set of projects we've restructured the way NQF evaluates measures that are already endorsed by NQF and are going through their maintenance of endorsement review, otherwise known as maintenance measures. That's the shortcut. And once NQF endorsement is granted, we do anticipate re-reviewing measures approximately every three years for a maintenance of endorsement evaluation.

Things change, but more importantly the world around us changes. The needs of the marketplace change. And so this is such a dynamic environment that it's important to be

sure that the measures continue to meet the criteria if they're NQF-endorsed.

Now what we've done in the past was always treat maintenance measures and new measures identically as if there was no history to the maintenance measure, which was sort of dumb. And hearing that feedback, it's like, okay, yes, that makes a lot of sense, particularly since we house all of the information from the previous evaluation in our data system. We've got it. So it's a matter of just pulling it up and looking at it.

So I think for those of you who might be working with us with your maintenance measures right now you'll find that what we're asking you to do is just look at the we already have and update it. Don't replace it. Just update it. And that way we have the historical view of what happened last time, you don't have to rework or rewrite anything. Yes, we've changed a few forms. It's okay. It's all about the same stuff. Use the old forms. You don't have to

redo on the new forms. It's fine. And realize that we're focusing in on things that are really likely to change over time about a measure.

Yes, evidence can change occasionally now and again, but mostly it doesn't. And usually what we're seeing is the evidence -- if anything else, if there's something new, it's additive. It just adds something new and says, yes, it's still good. It's rare that it changes it dramatically. Sure, there are isolated examples, but that doesn't happen in most measures. And so, there's not a lot there.

Similarly, testing for liability and validity, if you've done it well the first time, you may do it again. And it's nice to see updated data, but it probably will not be the decision point of the measure. What's really important on a maintenance measure is what's happening. What's the current performance? What are the results over time? How many folks are using the measure? How broadly applicable is it? In other words, what do we know about how good it

is as a tool to drive quality improvement? How's it working? What are the problems with it?

So the focus for the maintenance evaluation is on current performance, opportunity for improvement and what have we learned by use of the measure, really focusing in on use and usability. Feedback from whoever. It might be the MAP. It might be folks in the field. It might be feedback from our committee members. But we really want to know if it's working. What are the issues? What's happening out in the real world outside the Beltway? So maintenance has really changed the focus of looking at those measures for ongoing endorsement.

To help facilitate the logistics of that, we created the maintenance checklist, and with all of your help we've gotten through the rocky early stages back and forth. We've clarified some things. We've listened to your feedback and hopefully I think we're in a reasonably good place right now. But the checklist is nothing more than the directions to

you all about what you need to do to update it for maintenance.

And for a lot of measures there's very little. We've had to really emphasize that in order for us to see those changes you need to write them in in red, otherwise they just blend into the background and we don't know where the new stuff is. That was something we had to learn together. But if stuff hasn't changed, all you have to do is say it didn't change. Nothing's changed. Same stuff. We're good.

And so, just want to let you know it was clunky to have to use the add-on extra SharePoint site to do the checklist, but our plan is to incorporate that within the whole OPUS system with some skip logic and stuff like that hopefully in the coming year. So this was a temporary fix and we really appreciate your indulgence in working with us, but we find the information when you tell us what's happening with your measure with these quick little answers and following the directions on the checklist

limits the amount of work you have to do to really get that measure ready for submission.

So at this point though we've done several projects. We've got a lot of submission deadlines coming in. Any questions, comments about the change in this process that you're experiencing?

(No audible response.)

DR. WINKLER: We've been getting feedback on our monthly calls. We get feedback from our developer Advisory Panel.

It's your turn. Anybody, comment? Jeff, are you saying something?

MEMBER GEPPERT: So what can you say about sort of the presumption of -- for endorsement, for continued endorsement? Is there an explicit or implicit presumption that in the absence of positive information?

DR. WINKLER: Yes, I think that one of the things that's happening is I don't think you can have that presumption, because the environment around us changes. The marketplace

changes. The demands of potential end users are changing. And so, I think that any time a measure is evaluated it has to be done within the context of the environment of the time. And it may be quite different than three years ago. And that's the reality of this very dynamic measurement world.

So I don't think you can say there's a presumption. There is and we have said this, that there is a general movement towards outcome measures, composite measures, patient-reported outcome measures, less so on basic simple process measures. And that evolution we're seeing and we're seeing that as changeover within the NQF portfolio of endorsed measures.

So again, it's the environment and the marketplace in terms of the kind of measures that are proving themselves to be particularly useful and valuable out there.

MEMBER GEPPERT: Just I think in terms of the evaluation of the measure I think it would be helpful to be explicit about what those

environmental context changes are that motivate the decision.

DR. WINKLER: Point taken.

Question?

MEMBER BLAKEY: Hi everyone. My name is Alicia Blakey. I am representing the American College of Radiology. Just wanted to thank you for your presentation on just the maintenance of previously endorsed measures. I'm kind of new to the college; only been there for probably two years, and just recently had to go through three applications kind of back to back.

So I think just in general because I don't feel like the process for filling out the application was as easy as it's sounding today.

(Laughter.)

MEMBER BLAKEY: You know the measures have been incorporated in accountability programs. Not a lot has changed. Of course performance is improving. And I just wanted to get just more information on kind of what -- the forms online it requires you to kind of redo

everything again. It's not necessarily just check yes. There's no -- so for example, let me give you a practical example.

One of our measures has been a program since 2000 to 2007, 2008. And we did not submit any additional -- we submitted additional evidence that was not necessarily systematic reviews, but really talked to the use of the measure and its importance in filling a gap in care. And the committee, their preliminary analysis said that that was insufficient evidence.

DR. WINKLER: Yes.

MEMBER BLAKEY: And so you can imagine if your evidence is not graded, then I think that was the rationale, but something as simple as we have great evidence for it was not sufficient.

DR. WINKLER: Yes.

MEMBER BLAKEY: And just even I guess in general I just wanted to make a comment. If you could explain a little bit more in depth of the training for each of the project teams. So

some project teams it seems to be inconsistent how they -- it seems very subjective as far as how much intent they give to evidence or reliability or in validity. So it's not really consistent from -- like we're working on two different projects at once, so I see a complete difference.

DR. WINKLER: Okay. And we're pleased to hear your feedback because that's something we're actually working exceptionally hard with. As Sarah and Karen and I know, we put at least two to three of us doing those evaluations to try and foster that consistency. But again, it's a work in progress.

A couple of things: When you're looking and answering the questions, it's important that you understand what the criteria are, and the criteria are very specific. And that's what we're using to guide the committee to do the evaluation. So for instance on your example about evidence, the evidence for a measure is expected to be a systematic review of

the body of evidence that relates what you're measuring to patient outcomes. And that's what we're looking at.

And so, it's important that you really take a look at the guidance document that really explains the criteria and how those will be evaluated to understand. But again, we're happy to help you out with a little bit of off-line technical assistance any time if you do have questions. No problem with that.

MEMBER BLAKEY: Thank you. Yes, I think the technical assistance will be helpful just because the process should have been a little bit easier than I had anticipated. For new staff who are new to measure development, it's good. I think technical assistance is helpful.

DR. WINKLER: Yes, I think for new staff and new to measure development it's like coming up on the on ramp on a freeway. It's moving fast, folks.

Question?

MEMBER KAYE: Hi, this is Toni with the PCPI. I wondered if you could speak a little to have you changed the way you kind of advise -- have your committees and kind of train them to the focus for measure maintenance?

DR. WINKLER: Absolutely.

MEMBER KAYE: I'm thinking -- first example, whenever this was first rolling out and the idea of if the evidence hasn't changed or if it's only been added to, them don't worry about it. But I think in practice sometimes we've heard with committees it's more a, well, do you want to review it? So there's a chance of kind of being caught like, oh, we didn't know they could even discuss it. So it's kind of a matter of how you prepare going into the meeting.

DR. WINKLER: Right. I think that ultimately since the committees have the final say we do -- our guidance to them is, hey, guys, if nothing's changed, let's not waste our time here saying -- doing -- redoing work and let's move on. But the committee always has -- it --

because it's ultimately their responsibility, they always have the opportunity to jump in.

But we have found that; at least my personal experience is, they're fairly guidable, mainly because there tends to be a lot of work to do. And if there's something they can like limit and move onto something else, they're pretty good at doing that. But it's we're trying hard to guide them there, but it's not like we can totally, totally force them there because we have granted them the role of the oversight and the responsibility.

So my experience though is that they're reasonably happy to be guided. I can't promise that's going to happen all the time. Perhaps there may be some controversial measures or controversial areas, but by and large they're looking to focus a lot on talking about how a measure is being used. What happened at my house. Or they want to tell you their horror story or their -- what's going on.

So, but again, I think it's a work in

progress. We're learning. We share among all of us who are doing this and leading these projects. We're in constant contact. Even during a meeting I had Karen emailing me, how's it going? What's happening? Did this work out? Did that work out? So I mean, it's -- we're making all the efforts we can to use a common approach and a common experience across all the teams. But it's challenging because it's as new for us as it is for all of you. So thank you for all of your help, assistance and indulgence in doing it.

MEMBER KAYE: Thanks.

DR. WINKLER: Okay. Anything else? I'm not seeing anybody.

(No audible response.)

DR. WINKLER: Okay. Next slide. Off-cycle. I think I hand this one off to Sarah.

MS. SAMPSEL: And I'm just going to -- I was trying to figure out if I wanted to say this or not in response to your question, Toni; and maybe if I haven't had enough coffee, I shouldn't say it, but I echo what Reva says, that

it's a learning process for us, too. I think evidence is not the hard part. It's when -- I think the testing updates. And so, as many of you know who have been around for awhile, sometimes you get an original form back that's the old NQF forms, and you know they look very different from the new NQF forms, which reflects updated criteria.

And so, what we're trying to do on our projects as well is get ahead of that a little bit, not only go back for the history, as Reva said, which is why we want to see that documentation of what happened beforehand, but then also being able to say, okay, they've updated their testing.

So the maintenance checklist said they updated their testing. Well, then we want to know why did they update their testing? What was that change? And sometimes those old forms don't reflect it very well. But that's where we're trying to get ahead and have the conversation with you before we go to committee so that we can

help represent as well what that change was. But again, sometimes that comes down to the technical assistance. To me this is all woven together.

What I will also echo what Reva said from her earlier presentation is new developers, some of these novel new measures coming out, technical assistance is key. And we really are seeing, I mean, I really have to say beautiful submissions. And that's hard to say, and that's really geeky, too.

(Laughter.)

MS. SAMPSEL: But some just really nice submissions. I go they listened to us. They listened to us. And you're really happy about it. And I think we all feel that's a success because it helps us.

And we also as senior directors kind of throughout the teams we want to know your measures. We want to help you. We want to be your partner in this and it's hard to do if you don't come to us in the first place. And when you get a call 24 hours before a measure

submission saying, yes, I don't know how to fill out this evidence question. I don't know if this is a process measure. Well, it's really hard for me to want to help you. I mean, it really is. So, anyways. But we still try and it's a give and take and we understand that.

So off-cycle activities. One of the things that we've been working on, and probably a little bit slower than we wanted, but you're going to start seeing beefing up over the next couple of quarters are off-cycle activities. And so, we started using standing committees a couple years ago now, therefore we have a committee that's assigned to a content area that's responsible for the portfolio over time. And so, and they're responsible for the full portfolio, so you should have the same standing committee if you come to us this year, next year, whatever.

There are sometimes little tweaks to the standing committees, but what that also means to us is we need to find a way to keep them engaged. We don't want to just call them every

two years and say, hey, you have -- you need to look at these measures again, or we have some new measures for you. We want to keep then ingrained in the process.

Again, to go back to things like when we make updates to the standards, this is an opportunity to educate. Hey, we know you're not active right now. We know you don't have any measures to look at right now, but we've updated our maintenance process, we've updated X process and we want to keep you informed.

But we also want to give them an opportunity to remain aligned and engaged with each other. And so, we're looking for opportunities in working with our standing committees. What do they want to talk about off-cycle? You don't have 25 measures to review this year. What do you want to talk about? What can we do to help you stay up to date and current, not only in NQF, but what's going on in the measurement field in your area?

Next slide. Where did the clicker go?

I guess I'm in charge of next slide. Whoops.

So basically this is a kind of laundry list of some of the things that we're thinking about with off-cycle activities. Serving as a clinical or a technical expert panel for other standing bodies. So it could be Measures Application Partnership. It could be one of our white papers or other workgroups. It could be kind of something new coming out, whether it's in the quality innovations department, whether it's the incubator, some expert panel to bounce ideas off of of what's going on.

I'll give you an example here in that we have a number of measures that were submitted to Person and Family-Centered care that really weren't measures, they were tools. So we had a conversation with them and said we really can't send these through, but we're willing to give you an hour on an agenda. We have a seated committee right now and the standing committee, and they're going to meet. We're going to give them time on that agenda.

But I would say in an off-cycle year, an off-cycle period we would be able to convene our standing committee and say, hey, we want to talk about this, or, hey, we have somebody who's really interested in creating a measure on some kind of PROM or some kind of shared decision making or something like that. This is an opportunity for you to give early feedback and help guide this discussion. So those are the types of things that we're starting to think about with off-cycle activities.

Connecting the standing committee with external entities to discuss and identify potential collaborative groups. This might be the case, and again we're probably going to be doing this in the renal project where we have a process measure based on a PRO-PM. There's really nothing wrong with a process measure, but they really want to be going towards outcomes. They have no idea how to do it. So why not put the experts -- put them in front of the experts and have the experts give them some of that

feedback? And we can use that. It's not measure development advice, but more about these are some concepts you can think about. This is what we'll want to see when you come back for endorsement.

Sharing of innovative performance measurement work done by committees. So as you all know and we talked about a little bit yesterday are gaps and identifying gaps. One of the purposes of that conversation is to also find out what's going on and what is coming down the pike for future activities. This is an opportunity again to share in specific ideas and examples of those innovative ideas that are coming out.

I think this next bullet is pretty close to the last couple that I mentioned. Educational activities and then ad hoc measure review. So ad hoc measure review, some of you I know who are here today and some of you may experience in the past or in the future, if you make a material change to a measure during your annual update process, you're going to be

notified that the measure is now going to be scheduled for an ad hoc review. We will be doing some of these during the off-cycle reviews.

So let's say behavioral health we only have one measure scheduled for review in a given year, but there are a number of measures that have been identified for ad hoc review. They would go through an off-cycle committee versus pulling the whole committee together for a two-day meeting, etcetera.

But we would be giving you -- and there will be a back and forth exchange of we'd like to do it in fourth quarter. You'll hear from staff. Will you be able to meet that guideline? And we should also be giving you feedback on why we feel that your measure needs to have an ad hoc review because of some kind of material change. But we would also hope that if you've made the change and notified of that you were already aware of what you did.

Next slide. So this is just an example of one of the -- we just finished up an

off-cycle review, and it was the Care Coordination Standing Committee. And mind you, some of our committees the experts really just like to be engaged a lot. And so they're always looking for ideas of why can we meet, how to stay engaged, and we're happy to kind of continue that conversation.

But this was kind of an ideal situation that when there was outreach both to CMS and ONC, as well as to the co-chairs of the standing committee, they were just like, hey, we have this great idea. Let's talk about current activities around health information technology and how they relate to care coordination.

So basically we convened the Care Coordination Standing Committee via phone. Or no, they -- did they meet in person?

Okay. It was a webinar where they had a conversation about what ONC wanted to talk about and they were able to release a report. It's not as thorough. It's not a 300-page report like we get out of a full cycle, but to the same

degree it was the opportunity to facilitate a discussion that was generated by ONC using an expert panel and coming up with that idea of what they need to -- what they would want to talk about.

So upcoming off-cycle work that -- so you should be aware. Behavioral Health Standing Committee will be doing outreach to our co-chairs in the -- well, it was supposed to happen last week; it didn't happen -- next week regarding what are potential topics to the co-chairs? And we also need to assess the standing committee membership because we have some folks who are -- their terms are up.

But basically, talk to them. And I'll be reaching out to say what are those things that are going on in behavioral health that you might want to talk about that is an opportunity to engage and hear from the standing committee members?

Kind of one example that we're seeing a lot of is the integration of behavioral health

in chronic conditions. What's going on in measurement in primary care? What's going on with diabetes depression measures? What's going on with some of those other chronic conditions and how they affect some of the behavioral health measures. So that might be one opportunity, but frankly, I don't know. We'll reach out to Peter Briss and Harold Pincus and get their feedback on what they think this group will want to talk about.

We would anticipate pulling together that webinar in the third quarter. That will be released out on the -- on our web -- on the project page. And then also we'll be doing fourth quarter ad hoc reviews. So for those measures that have been identified with material changes during their annual updates or perhaps they were measures that had conditions of endorsement during their last phase of work. Those will come back for ad hoc review. And we're teeing that up to happen in the fourth quarter of this year.

So I am staffing that as well as Melinda Murphy because I'm conflicted with most behavioral health measures. And I'll just say that publicly. Not that I have anything against them, but I worked on most of them. So it doesn't make sense that I can also review them at this time.

Similar situation with musculoskeletal, that we'll be talking to the co-chairs, looking at the standing committee membership and coming up with some ideas to reconvene musculoskeletal, because that group hasn't met for a couple years. And I'm similarly conflicted with musculoskeletal.

So I am just really conflicted, but I keep coming up with a real laundry list of what I'm conflicted with. Some of it's flying; some of it's not.

So that is that. I guess what I want to do: one, see if there's any questions; and two, I really want to know for any of those of you who have been indicated for ad hoc review,

are we giving you enough information? Again, this is your opportunity to give us feedback. Where can we do a better job and what would you want to see during the off-cycle process and some ideas on how to engage. And maybe some of you have some ideas for some of these committees, what we could be talking about.

Yes, Sam?

MEMBER SIMON: So in the vein of be careful what you wish for, here's one of those questions about ad hoc reviews. No, it's fine.

(Laughter.)

MEMBER SIMON: No, just generally speaking, I mean, I know -- I understand what the purpose of an ad hoc review is. There's a change made to a measure and the committee wants to understand sort of what change that -- material effect that has on the measure. But just like Reva gave us some broad high-level guidance about what the committee looks for, what types of evidence the committee looks in a maintenance review, I think the same would be really helpful

for an ad hoc review. Are you looking for performance data? Are you looking for the measure developer to go out and collect additional data about usability and use? I think having some high-level guidance along those lines would be really helpful.

MS. SAMPSEL: So the easy answer to that is technically it would be the same as a maintenance review. At the high level you're still looking for usability, but you're also looking for -- and this might be an interesting exercise of what is a material change and -- do you have something to say?

DR. WINKLER: I'm sorry.

MS. SAMPSEL: I'm just going to let Reva go.

DR. WINKLER: Only because I got to dance with the CSAC about what is a material change. And the most recent guidance that's posted on our web site is the most recent wording they've come to conclude. And so, a material change is, as everybody says, you know it when

you see it. Well, good luck with that one. But, so I would refer you to the wording because I don't have it quite off the top of my head.

But when we do identify a material change, one of the things CSAC did say is they would expect to see new results using the measure with the new specifications. So they do want to see some data associated with it.

That would be the one thing with material change are the specifications that they are looking for is because if - presumably it's material, it's going to change things. The question is how much and how significant might that be? So that would be sort of the in addition to just the general criteria is because it's new spec you'd expect some new data to go along with it.

Question from anybody else? Yes, Kyle?

MEMBER CAMPBELL: So regarding behavioral health we're currently developing measures for CMS in the inpatient psychiatric

facility setting, and we have some measures in the pipeline, but our understanding is there isn't yet a date for a behavioral health CDP. So I didn't know if there's any opportunity -- like if you had any guidance about when that might happen. And I think we would probably have three measures ready by the fourth quarter.

MS. MUNTHALI: Fourth quarter you said? That's good timing. We can't confirm anything right now, but if you have it by fourth quarter of 2016, that's good timing.

MEMBER CAMPBELL: Great. Thanks.

MS. SAMPSEL: And I do want to say more globally, too, is remember that's -- the issue with the standing committee is the ability to always keep us informed of new developed -- of new measures in development. And so, if we know about it; and now we know, is then that opportunity to have conversations with CMS, but then also to think about integrating to off-cycle, because we may be able to -- as an example, when I was talking about the PRO-PMs

yesterday, the patient activation measure, that was reviewed. That's a brand new measure reviewed in an off cycle.

So, there might be some opportunities there where we're getting them together. And if it's a synergy, they may not come together in person, but we might be able to do that review.

MEMBER CAMPBELL: Sorry, I have one more follow-up question. How does the integration work? So we submitted a 30-day all cause unplanned readmission measure for the IPF setting for the Readmission Standing Committee to review. But is there collaboration between the two committees, because that's something that the Behavioral Health Committee was interested in seeing developed for that setting.

MS. SAMPSEL: So I think there we probably have a lot of case studies more than anything else in that we all stay in tune on what measures we're reviewing and what are coming through our committee. And as an example, Karen's Managing Palliative Care and she has pain

measures. I'm doing pain measures in Person and Family-Centered Care. So we're trading off reviews of the measures and making sure that our committees are aware of what's going on. So that's a really good example of now we have a trigger of I need to follow up with Readmissions to get that feedback from Behavioral Health.

Some of it's a little bit informal where what we might do is -- and here's another example. I have a measure, a home and community based services experience of care measure coming through. I've noted -- I've given the Duals team and the HCBS team notification to send that out to their committees to do some of the public comment on those reviews.

So there's not always formal integration because we don't want to totally decimate one standing committee over the other, but we do know that they have feedback and we do want to get that feedback.

MEMBER DORSEY: I just wanted to ask one question, a point of clarification around new

measures. So I think I heard you say that it's useful for you guys to hear when we have a new measure way ahead of when projects get scheduled on a calendar. And then you guys can work with developers to figure out when --

MS. MUNTHALI: Yes, it's really helpful for us to know that because it feeds into our negotiation process with CMS. And that oftentimes helps us to prioritize which projects we're going to have.

DR. WINKLER: Yes, I mean, we can't emphasize enough that what we know that's coming down the pipeline is really important for all the planning, because you know there is a lag by the time you organize the funding, create the projects, get things going. So the more up front we know when what's coming and when it's going to be ready, the better we can plan to incorporate it into our work.

So again, you don't have to wait for us to ask. Please tell us. That would be very, very helpful and I think it works out well for

all of us.

MS. SAMPSEL: What I'll just add to that is remember in all of our project pages there should be a project email box. That's all you have to do is send an email to the project email box.

MEMBER KEENAN: Do you all look at the MUC list to see measures that are coming? Oh.

DR. WINKLER: Hold that thought.

MEMBER KEENAN: Okay.

(Laughter.)

DR. WINKLER: Hold that thought, because, yes, and that's going to be one of the important aspects of the CDP-MAP integration we're going to talk about a little bit later.

Okay. Any other questions about the off-cycle work? We're going to go into one more subject.

(No audible response.)

DR. WINKLER: Okay. And that's just a quick review of something that's changed this year and will go into place in the last half of

this year. The specific date we will announce when we have it finalized.

But just to -- remember, most of your effort is always on the -- up front of the CDP steps, but the back end, the later steps are also very important. And for those of you who are familiar with it, you know that the endorsement process is essentially the committee makes a recommendation, it goes out for comment, goes out for vote, and then those recommendations go to the CSAC, who makes an endorsement decision and the board of directors ratifies that endorsement decision to grant the endorsement.

Appeals, there's the sort of final step required within the CDP process so that an appeal can be submitted within 30 days after the endorsement is announced. And those appeals have gone to the CSAC for review and then again to the board of directors for review.

That process was reevaluated. And the question is why make any changes? Well, we undergo a lot of scrutiny about our process, and

this one actually involved the board of directors. So the board started doing a little self-scrutiny and thinking that it needed to be reassessed, particularly to become more efficient and eliminate some redundant decision making and prevent redoing and re-discussing the same thing too many times and then reinforcing some of the finality of decisions.

And so, the board came up with some process improvements. And the principles behind them is that the CSAC now will really -- because they always have been the group that's able to really look at what happened during the consensus development process, that is their charge. That is what they do, unlike the board of directors, which has all sorts of other governing responsibilities, and this isn't their main job. So the CSAC it is their main job. And to separate the role of the CSAC when it comes to appeals, a newly created appeals board will decide measure appeals. And so, appeals of the endorsement decision will not go back to the

CSAC, but will go to the appeals board.

So again, so pictorially; I think I've got -- yes. The process, if you notice we've got one less arrow on each sign, so that's more efficient. So essentially the recommendation from the committee to the CSAC after voting and comment doesn't change. Same stuff. But the CSAC's decision is the final decision. The endorsement is granted at that time. You'll find it happens two to four weeks earlier than the previous process. So it does shorten the whole process overall.

The appeals process then. CSAC is sort of the starting point granting the endorsement. If an appeal is submitted, then it is presented to the appeals board for review so we don't have that multiple couple of extra steps back and forthing, and simplified the process considerably.

Now the grounds for appeal were also revisited. And so, if you noticed that the old grounds for appeal were to be filed in response

to any of the endorsement decisions and it needed to include some written evidence as to why do you care and what are the issues? And so, but in the new appeals process there are two reasons for appeal, and one is around procedural errors such as failing to follow the process as written out, or there is new evidence or information unavailable at the time the original endorsement decision was made, right, that wasn't considered that's reasonably likely to affect the outcome of the original endorsement. So that's a relatively straightforward change to the appeals process.

The appeals board will consist of current board members, former CSAC members and former standing committee members. So nobody that's actively in the process but has sufficient experience to understand the process and NQF's activities.

And so, what the appeals board will do is review the appeal and then make a decision that will either uphold the endorsement decision, overturn it or dismiss the appeal. So they

aren't re-litigating any of the criteria or anything like that.

So this process has been approved by the board. It will go into effect the second half of this year. The exact time frame of which project we're going to flip it on is to be determined and -- but I think you're going to see it in effect for the second half of 2016.

So any questions about that change? I mean, we've been talking about this a little bit, so you may have heard about it before, but we wanted to be sure everybody's familiar with it and aware of it.

Yes, Matt?

MEMBER POPOVICH: What was the percentage of measures over the past two years that were appealed? And then like is it a significant process or is this just -- and then where does the measure developer fit into this --

DR. WINKLER: Okay.

MEMBER POPOVICH: -- within the appeal decision making?

DR. WINKLER: Actually the number of appeals is relatively low. It's a handful over the last two years when we've done several hundred measures. So we're talking about a low percentage.

Where does a measure developer fit in? The measure developer is involved in the whole process at every step along the way. As you know, when we take the measures and the results of comment and voting and the committee recommendations to CSAC, we -- measure developers are -- hopefully attend that and may have an opportunity to contribute to that conversation, depending on what it is.

Again, with an appeal the folks we notify that an appeal has come in is the developer, an opportunity to respond to what's going on. Often -- and it's particularly pertinent if there's something technical about the appeal that's about the measure specs itself. The developer is often the crucial responder about what -- a decision is made or not made

about how the measure is constructed or something like that. You will be less of a player if the appeal is about a process or something like that.

So, but definitely you will be involved in the whole thing and you'll be notified if an appeal is submitted and that we will be taking it to the appeals board. And you will be notified when that happens and the whole thing. It's just another lovely step of the process we all get to share.

Any other questions? Question, somebody? Jeff? Oh, okay. I saw something going on over here.

MEMBER GEPPERT: So in terms of the criteria; could you go back to the criteria, I'm just wondering what new and unavailable means?

DR. WINKLER: How a --

MEMBER GEPPERT: So the difference being new information that has recently been generated --

DR. WINKLER: Yes.

MEMBER GEPPERT: -- which would

probably feed into like an ad hoc process, or just information that was not made available or was overlooked.

DR. WINKLER: I would guess it's either.

DR. WILSON: No, it's new evidence.

DR. WINKLER: Okay. I'm going to give this one to Marcia and let her clarify.

DR. WILSON: I think the intent was that it is new evidence that comes to light, because you figure by the time a measure gets -- is brought forth to NQF and goes through the process some time has elapsed. So it's not like, oh, we forgot to say this during the original process when that measure is being seen. It is truly for new evidence that would have come to light.

MEMBER DORSEY: So that's really clear and I get it, but what -- how do NQF and measure developers -- how do you intend to handle when appeal letters don't necessarily follow these guidelines, right? Because sometimes these

letters the kitchen sink can get thrown into the content of what's brought up. And so, then what's the expectation for response?

DR. WILSON: One of the reasons for changing the grounds for appeal was exactly that. We felt that some of the appeals that had come in historically were if I say this a different way, I can appeal this. So that was one of the reasons for changing the grounds.

The other thing that Reva mentioned is that any appeal that comes into NQF will go to the appeals board and they can determine whether that appeal meets the appropriate grounds. So you may -- someone may file an appeal. It will -- staff will look at it. They will pass it to the appeals board and the appeals board will say, no, this doesn't meet the grounds for appeal and then the appellant would be notified.

DR. WINKLER: Anything else on that?

(No audible response.)

DR. WINKLER: Great. Okay. Isn't my turn over? Yes, right. But my turn's over,

right?

MS. ISIJOLA: Yes.

DR. WINKLER: Okay.

MS. ISIJOLA: So one of the things we wanted to also shed some light on is our IT space. Many of the ways that you contact or interact with us is within your submission dashboard, but also one of the things we're doing is revamping that and reinventing that. So Jason Johnson is actually our Vice President of Information Technology and he'll go over some of the changes that we'll be making and that you'll see within your dashboard within the next few weeks. But also this is an opportunity -- if there are improvements or suggestions that you may have that we can continue to reinvent the wheel as we make it easier for you to interact with us, he's your guy.

MR. JOHNSON: Thanks, Wunmi.

Hi, I am Jason Johnson. I'm Vice President of IT. I have been in that role only for about three months, but I've been here at NQF

for about five years. And there is a specific change that we wanted to talk with you about first, and it comes from the inclusion of eMeasures more frequently in our work. And the fact that we're in a bit of a transition period where our eMeasures -- we have eMeasures coming in while we have our regular measures, or what we've always called measures there as well.

And so, what we've done is made an accommodation where -- that we hope is clear to you and also to our outside users, particularly those people who are using QPS. Because what we have now is two different formats of a measure, what we traditionally called a measure and an eMeasure being referenced under the same number.

So we may have Measure 0070, which is one I just mocked up for an example. And if we use that single number, how do we differentiate between the two formats that we might have, the eMeasure format and the format everyone has been used for years and years.

So what we've done is tried to clarify

that by adding another number. So your eMeasure when it comes in will also be assigned a number. And what you're thinking of as your traditional measure number, or traditional measure will also be assigned a number. So they'll both have two separate numbers, but they'll exist under the regular measure number.

So a picture is worth a thousand words and you should be able to see that a little more clearly here. So here we have Measure 0070. And this is what the QPS display will look like. And if the user clicks on the little plus up next to it, they'll see that the measure is available in multiple formats. And then there will be two additional measure numbers, or rather NQF numbers available and assigned to that measure.

Now end users will be able to reference either of those numbers. So if they search by 0070, they will get both formats. If they search by 2906 in this example, they'll see all available formats as well. So we want to be able to present all of those to the users but

still allow the individual submissions; the eMeasure submission and the regular submission, be referenced by a single unique number. And that was the goal of what we were trying to do with this system.

When we display this more generally, you will see that we concatenate the numbers. So it will always be the first number that we're used to, 0070, and then there will be a colon and then the additional format number, 2906. And we really do expect this to be a transitional approach. Overall we'd like to keep the numbers that you have to reference and manage manageable and as few as possible, but we think that this display works best given the options available for our end users. And we're also trying to make it easier for you.

If you can advance to the next slide. So within your dashboard this is what you will see. The two different formats will roll up underneath that same measure that you're used to from your prior measure number.

So I'm going to pause there and see if there are any questions about that or how that lays out or why we made that decision.

Yes?

MEMBER CAMPBELL: So one of the things that I think would be important to consider -- and I like this that they're linked together, but I think part of the confusion in the marketplace with regard to alignment relates a lot to like different NQF numbers being assigned. So someone sees NQF 2906 and to them NQF 2907 isn't the same thing. So I'm wondering if there would be an opportunity to make it NQF 0070A, NQF 0070B.

MR. JOHNSON: Yes, we did consider that. So our intention in the marketplace is that these will always be displayed together. It will always be 0070:2906, so that we can that reference point. Or it would be 0070 alone, referencing the entire -- both formats.

We intentionally avoided adding extensions. We considered adding an E to the end, considered A, B, C for multiple formats.

And what that begins to do is cause problems for some of our end users who are used to downloading this and working with it in Excel. As you begin to mix alphanumerics it can begin to cause some sorting issues for folks. And we didn't find that quite as useable. We also have a number of internal processes that are heavily Excel-dependent and we thought that would make it a little more troublesome for us as well. So that's in terms of adding just alphas.

In terms of adding an extension like an E or something like that, we were concerned about overloading the number with too much meaning. We wanted the numbers to stay -- from a purely data perspective it's not wise to begin adding meaning into the number if you can help it. So we intentionally shied away from that as well.

MEMBER CAMPBELL: So one thing then I would encourage, because I think it sounds like from an internal perspective it works really well, but from a patient or a stakeholder

perspective when these measures get publicly reported by CMS and you have to go and the patient has to see NQF 0070:0296, 00 -- whatever it is, I think part of the problem that we're having in the marketplace and patients and stakeholders are understanding measures is a confusing numbering system.

MR. JOHNSON: Well, we hope that 0070 is the way that most are going to reference it and that the different formats are really more of an inside baseball-kind of piece, but I totally understand that approach.

The other thing that we believe is that this will go away, that this is a transitional issue and that over time each of these will only have a single format again, and that would be the eMeasure format.

MEMBER CAMPBELL: But I think it does go beyond the eMeasure format and that you might in the future -- as measures become more aligned across settings and across the payers you may have a situation where you have a physician group

measure that's 0070 and then you have a plan level measure. And I think the more that you can align that across the marketplace and across the settings of care so they roll up to -- because I mean, if you think about it, the underlying evidence for a hemoglobin A1c measure, regardless of what layer you report it at, is the same, right? So it's really just the testing results and the adaptation to a data source that's different.

And what we're seeing -- we're doing the National Impact Assessment for CMS. What we're seeing is there's a lot of confusion because a measure in a given reporting program has a completely different NQF number even though conceptually it's very similar. So just something to think about beyond the eMeasures as you guys sort of build this, because I know the behind-the-scenes part is hard and how it all links to all the systems you have.

MR. JOHNSON: Okay. No, thank you. That's very helpful. Any other questions about

this part?

Yes, in the back?

MEMBER McKIERNAN: So can you talk a little bit about what will happen when measures are retired or if one measure -- either like the paper-based measure or the eMeasure lost endorsement how that process would be handled? If the eMeasure was maintained, for example, would you continue to use both numbers or would it revert back to the original number?

MR. JOHNSON: It would revert back to the original number under the current plan.

MEMBER McKIERNAN: Okay.

MR. JOHNSON: The reasoning behind that is there's a lot of history behind that number and we want to retain that history. And we think that's going to be the clearest for outside end users, maintaining that continuum. Because again, we're really thinking about this NQF less as a measure number and more as a -- the NQF number associated with that format of that measure.

MEMBER POPE: Can you explain a little bit about how you came up with 2906 and 2907? Is there meaning behind that?

MR. JOHNSON: There is none, and that's intentional.

MEMBER POPE: So --

MR. JOHNSON: It's -- they're largely sequential.

MEMBER POPE: If that just gets separated, you wouldn't know that's related to 0070?

MR. JOHNSON: Correct.

MEMBER POPE: But you're not planning for it to get separated, I guess.

MR. JOHNSON: Well, we're not planning for it to be separated and we're also making it referential by that number. So if you were to come to QPS and search for 2906, it will bring up 0070 as well and tell you that there's that relationship.

MEMBER POPE: But in other instances they won't all start with 29, you're saying?

It's purely random?

MR. JOHNSON: Correct. It will -- we'll get into the 3,000s and 4,000s.

Any other questions about this part?

(No audible response.)

MR. JOHNSON: Okay. Well, I will explain a little bit more about some of the changes that have happened and that are coming to the dashboard. Some of you may have seen we added a developer section to the bottom of the dashboard where we can update and post materials specific to you.

But starting in August and through probably about October of this year our current intent is to begin working on changes to the measure submission form. And these changes will involve some changes to the taxonomy that we're using, but also improvements focused largely on trimming the length of that form, reducing the number of questions and making it a little easier to move through, either through in-line help or skip logic.

And so those are the two current areas of focus. And of course we're working closely with the QM Department on those. But the more feedback that you provide as you go through the process to the project teams, that will all filter up to us. And if you have any direct feedback, you can always send it to info@qualityforum.org. That always routes up to the IT folks as well.

So any questions about that? Yes?

MEMBER BERNS: Hi there. This is Samantha from Lewin. I think that adding in some sort of skip logic is a great idea, but I think it would be helpful to post in like a PDF or Word format also how that logic is applied, since I know a lot of us work on the forms off-line and then upload them into the portal. And so, it can be difficult to -- even when we do it now, we have a bunch of sections that are repeated then actually show up once. And so, I think that having some sort of user guide or information about how that's going to be flowing would be

really helpful.

MR. JOHNSON: Okay. Thank you. Very helpful.

Anything else?

MS. ISIJOLA: Are there any questions regarded to your dashboard? I know that's the way that you are going into your forms. Is the information on your dashboard accessible or informational for you? Would you like to see other things?

Is there anyone the phone who have comments or questions? Operator, can you open the lines, please?

OPERATOR: Yes, ma'am. At this time if you would like to make a comment, please press start then the number one.

(No audible response.)

OPERATOR: Okay. At this time there are no comments.

MS. ISIJOLA: Great. Again, info@qualityforum.org. We're trying to make it as easy for you as possible as you work through

your submission forms, as you gain the information. Again, was Reva mentioned, a lot of this information is on the submitting standards page. So if you can't find it, let us know. But as always, we do have our guide book that gives you all of this information about where to find things and how to find things. So let us know.

MR. JOHNSON: Thank you, everyone.

MS. ISIJOLA: Thank you, Jason.

Okay. Well, I think we're going into our next discussion item. We're going to talk about assessing validity, providing -- no, that was a test. We're actually going into our intended use segment.

You've been hearing about this for some time. It's becoming real. Our board has approved it. So Karen and Helen will be giving us an overview of what that means.

DR. BURSTIN: Good morning everybody again.

So we're going to talk about this whole intended use project we've done over the

last year or so and really what it's resulted in. And I'll tell you that the actual implementation of this is not going to be immediate, but there are some steps that will happen sooner than others that Karen will talk through with you. I'll go over more of the overview of why we did this, how this all began.

So the Consensus Taskforce was a group we convened for the board for awhile really trying to think about how do we handle what is often in committees a difficult issue, and certainly Karen Dorsey and some of our friends at Yale and others experience this a lot, which is it's very hard at times for committees, endorsement committees to look at a measure completely in isolation as if they don't understand the broader context in which the measure will be used, i.e., a measure may be used for a penalty program. Is it getting kind of a harder look than it might get if it was something perhaps used as part of more of a reporting program? So we want to be honest about this and

just have that discussion.

Now we can't sequester committees like juries. We can't put them away and say don't consider this. They all know it. It obviously exists. It comes up at all these tables. So one of the questions was should we think about having a more nuanced approach to our endorsement decision rather than a binary yes/no and potentially yes for this purpose, no for that purpose. So we had a recommendation to have a committee consider this issue whether we would endorse for intended uses or come up with different levels or grades of endorsement.

So ultimately we pulled together a really remarkable panel; I'll show you in the next slide, including numerous developers to help us think this through, to think about what are the various use cases for how NQF measures are used, thinking about whether there are enough distinguishing factors among the use cases, particularly around the science around reliability and validity and evidence, that you

might think about them differently.

And then think about the need for potentially how you look at different measure attributes depending on use. For example, would you expect a higher threshold for reliability and validity for a measure that might be used in payment was sort of one of the opening thoughts about this. And then think about whether we need to update our criteria and then think about what the pathway forward would be and how we would do this.

This was the panel that we had. So really just some great, great folks including several developers at the table, end users and as well as experts who have studied this and published on what do we know about the association of use of measures for reliability and validity. People like Andy Ryan, Beth McGlynn and Rachel Warner and others.

So here are the five recommendations from the panel that came out. We're going to -- Karen's going to hone in specifically on No. 2

for you, but ultimately what came through was that NQF should not try to distinguish between the measurement needs of these different programs and different accountability applications. And frankly, the science just isn't there to say this measure should have this level of reliability, this should have this depending on intended use.

They did say however; and this was sort of what we had worked through as one option moving forward, is that we would consider a new designation for endorsed measures that exceeds our current criteria and also includes a new requirement that Karen will talk more about around this idea of a requirement for vetting by those being measured and others.

They again reaffirm that our endorsement work should really be mainly focused on accountability applications. We talked with the panel about whether for example we should come up with a QI pathway and what those criteria -- less stringent criteria would look like. And the thought was there are thousands of those

measures out there. The world still needs NQF to really focus on accountability. So we abandoned that pathway for now.

And to specifically then encourage the MAP to think about how this new designation is, Endorsement+; we can't figure out a better name for it -- could be used as they think about the measurement selection program.

And the other thing that became very interesting, and this is this last recommendation about considering the interaction between the measures attributes and the program attributes, is there's probably some good measurement science to do there.

So for example, if you're looking at a measure in isolation at Endorsement Committee you don't actually know enough about how the program is structured and the way it will be used. Even if you say it's for payment, you don't know, for example, is it a threshold? Is it a percent improvement? Is it a ranking? So without knowing the program attributes, simply

applying a use case on the measure attributes in isolation didn't make sense.

Now with the MAP tables that's flipped a bit, because they were looking at the program attributes. And so, the question is how would they use this as part of their thinking? And we are going to try to pursue that additional work of thinking through how those two logically would come together.

Am I giving this to you now?

(No audible response.)

DR. BURSTIN: Okay. Karen's going to continue with what we're thinking of around this Endorsement+. Now is as good a time as any. Here you go.

MS. JOHNSON: Thank you. All right. So Endorsement+. So one of the things that if you want to give us ideas about a better name than Endorsement+, we're all ears. We've been through several and this is where we've landed so far.

So here are the criteria for achieving

this new label. So this would be applied to endorsed measures. It's just an extra label that would go on there. So it has to meet evidence for the measure focus without an exception. So it can't be -- it can't go through our endorsement process with the evidence exception on it. Okay?

Reliability has to be demonstrated at the score level. All right? The score level. Validity also has to be demonstrated at the score level and it can't be face validity. Right? So, and these three here we already have in our criteria. And to some extent it's pretty objective, right? We'll know if testing is done at the score level, that sort of thing.

The last one is the new one. The candidate measure is well-vetted in real world settings by those being measured and other users. So that's the new thing that the advisory panel thought was really important. And this is what our board approved.

So what we had to think about is what

do we do with this last one? And number one, how do you define what's well-vetted? Who are others? What's real world settings? Do we need to include this in our criteria? Those are some of the questions that we walked through.

And this is the punch line of the story, basically. We will be adding this vetting as an additional sub-criteria to the usability and use criterion. Okay? You guys know that at least right now that is not a must-pass criterion. We realize that this vetting business is as bit of a stretch. It's not necessarily something that everybody does. So it is our way to signal that our board and others think that's it's an important way to go forward in the development process, hence adding it to our criteria. And the definition of what we mean by vetting is here in these sub-sub-criteria.

So first, those being measured have been given performance results and data as well as assistance with interpreting the measure results and data. Those being measured and other

users have been given an opportunity to provide feedback on the performance and implementation. And then finally, the feedback has been considered when changes have been incorporated into the measure. We're still working on that language on that last one a little bit, but the idea is you don't just give people a chance to give feedback, but you actually use it to the extent that it could. Yes. Yes, you listen to them as Reva says. So this is where we are. When are we going to implement this? Probably later in the summer. When we put out changes to our submission forms you'll be seeing this. We'll be updating our criterion guidance documents, et cetera. You'll be seeing this.

So let me stop there and see if you have any questions. Yes?

DR. BURSTIN: So, thanks, Karen. Just one more thing to add. And we had a big debate about whether to add this new sub-criterion under validity, because you could also make the argument that part of the validity of a measure

is that those being measured and others think it actually measures what you hope it's measuring. But then it would be a must-pass. And so we're putting it here for now, I think, as really a signal, kind of a signal to the field saying people agree.

Really important, half of the debates we have around some of the measures being used in payment frankly are about the fact that those end users haven't really been engaged before they hit our tables. So some of this is really to think about how we get ahead of this storm, but we're not making it a must-pass. We're putting it out there as sort of a signal early on as we've done in the past around evidence and testing, but more so so that you can begin preparing to think about how you'll do this.

And we know this won't be easy for some, but would love your thoughts and suggestions. Is this too high a bar? Is this something you could work towards? And we'd just love your discussion.

I think Jeff had his hand up first. Did you have a comment? Okay. Go ahead. Yes.

MEMBER HIBAY: This is Sharon Hibay from Livanta. So I look at that -- the three criterion and each one of those could apply to an institution. So I think of when I was at NQF, just a network, a hospital, Boston Children's could submit this information. Is that appropriate? Is that defined as -- so when you think about what is vetted, it doesn't really give some sort of gradations or doesn't say in a network. It doesn't say on a state. It doesn't in some sort of national program, quality program. It doesn't say an HHS program. It doesn't say in a payer program.

So I could look at a Boston Children's measure; and we looked at some in cardiology, and I could say, yes, that meets the criteria as it's defined. So is that your intent that vetted could be at a network or a hospital level?

MS. JOHNSON: Sharon, the way I think about it is you're asking how much vetting do you

have to do? In other words, does it have to go out --

MEMBER HIBAY: Absolutely.

MS. JOHNSON: -- to everybody that's being measured or can you have like some testing data or something smaller? And that's one of the things that we've talked about. I think at least right now where we're at is -- and, Helen, correct me if I'm wrong. I'm thinking that Boston Children's would be a possibility. Just like anything else that the committees look at and think about that way, how wide was your scope? How reaching was it, that sort of thing. And they may or may not feel as happy with a Boston Children's as they would the whole nation.

MEMBER HIBAY: Absolutely.

DR. BURSTIN: I think we've left it intentionally somewhat vague.

MEMBER HIBAY: Okay.

DR. BURSTIN: I don't think going -- for example, if you're the developer of Boston Children's and you're going to just the docs of

Boston Children's, that's probably not really the spirit of what we're trying to get at here, but I don't think we're being prescriptive and saying you have to go out and do 100 hospitals or 100 end users. We also recognize that at least for the measures going through CMS for payment often times they'll have a full dry run that's supported and funded by CMS.

So we're not expecting you to be able to do a full dry run, but it would be nice to as part of this process to get input from those who will be measured in particular. We added and others, because I think others who would like to use the measure ultimately, like purchasers or regional collaboratives, would also welcome the opportunity to provide early feedback on some of these measures. So again, the idea is to just push the measurement earlier -- the feedback earlier and earlier in the process.

MEMBER GEPPERT: Yes, my second comment was related to that. Those being measured could be a pilot study, could be some

subset.

So my first question is about the validity testing. So there's multiple pathways beyond face validity. So one pathway is criterion validity, which could involve a chart abstraction for like a patient safety measure or comparing a risk model from claims versus a registry. And then there's a construct validity path. But do both of those pathways count?

MS. JOHNSON: Yes.

MEMBER GEPPERT: Okay.

MS. JOHNSON: It's not face validity.

MEMBER PANCHOLI: Hi, so I'm Mamatha Pancholi and I run the quality indicators at AHRQ. And I'm curious to get a more concrete sense of what we mean by giving the users their performance results in data when you're not CMS. So where is there -- so as AHRQ, I don't have hospital data. I'd have to go through a process of collaboration and getting folks on board with us. So are you thinking that as part of the submission process we would need to ahead of time

go through our process of providing folks with our specifications, having them run their own data and find out what their rates would be and submit that information back to us, so it's part of a feedback loop? Is that what we're thinking or is there -- or am I misinterpreting what you're --

DR. BURSTIN: Well, that's a literal translation, Mamatha.

MEMBER PANCHOLI: Right.

DR. BURSTIN: I mean, obviously that would be hard to do.

MEMBER PANCHOLI: Difficult.

DR. BURSTIN: And really difficult. I guess the question would be are there ways -- and I'll think about it in the AHRQ/PSI context because I think it's an easy one in some ways, because it doesn't fit the model as well, right? You have state data. You have state databases. But at the same time AHRQ does a lot of work with states and with hospitals. Could you have just some input from hospitals? Even as you've shared

results in past years, input you've heard from hospitals, how have you -- and particularly I would think with the PSIs you did a ton of this, in fact as part of our last Safety Committee, as you know very, very well around PSI 90 --

(Simultaneous speaking.)

MEMBER PANCHOLI: And we'll be doing it again soon.

DR. BURSTIN: But the key thing there was that that feedback was incorporated into the measure and the measure changed.

MEMBER PANCHOLI: Right, so if that's what you mean --

DR. BURSTIN: So you're kind of a poster child, I think, actually.

MEMBER PANCHOLI: Yes, so if that's what you mean, then that's fine. You're right, I'm interpreting it literally as --

DR. BURSTIN: Yes.

MEMBER PANCHOLI: -- a step forward, but it would be --

DR. BURSTIN: I mean, in some ways is

there some way as part of your development or a refining --

MEMBER PANCHOLI: Okay.

DR. BURSTIN: -- of the PSIs as you go forward share their information with a few states who might be able to get feedback from their state hospital associations, whatever the case may be, so you could indicate that before they hit our tables and you get the slam back from any at the table? Can you at least kind of try to head off some of those issues of the past, I guess?

MEMBER PANCHOLI: Right. Thank you.

DR. BURSTIN: Yes.

Yes, please?

MEMBER PEZZULLO: Lynn Pezzullo, PQI. So I think some of these questions are similar to what I'm trying to understand. I guess how structured or formalized of a process should this be? Is it enough to have an open public comment period that is distributed, the notice of that comment period is distributed specifically to

those users or those being measured, or should it be something more like convening a panel of those groups or individuals? So I just want to make sure that if we're considering this we're thinking through how do we really meet this criteria?

DR. BURSTIN: I think the answer is still kind of fuzzy, to be honest. This is brand new. That's why we're bringing it to you for your thoughts.

My personal sense would be that simply doing it as part of what you would already do as part of your comment period -- I don't think would necessarily satisfy this. You won't necessarily have as part of your comment period -- again, I could be wrong. I mean, I don't know how many for example pharmacists are responding back on PQA comments. So I think the question would be can you also build in -- and it doesn't have to be a separate panel, but is there some way as part of the process to begin doing something that's saying direct to those being

measured, I would say, in particular and say can you provide your feedback on this as we're moving through this process?

It may be later in the process that you might put out for public comment, but perhaps earlier than it hits our tables. And again, we're not making this a must-pass, but it's just something to at least begin to see the kind of information we get in. And I think what we'll do over time with you is try to be more clear. But this point we're leaving it open to your ideas and suggestions so we can see what comes in.

We're very much trying to handle -- I think at times we're just -- the measures that hit the tables where the end users are sitting and they create this kerfuffle that I think we'd love to have you kind of get a handle on before you hit those tables. And frankly, it may make a better measure by actually taking in some of those thoughts they have about issues that were raised.

And some of you may know the very long

process we had with the dialysis readmission measure last year, two years ago, Marcia, that Marcia sort of walked us through, in terms of bringing the appellants and CMS and the measure developers together. And at the end of the day there were a couple of pretty significant issues from the dialysis facilities that CMS was willing to accommodate, at least for the first couple of years of the measure being out there.

And so again, I think that early input from the dialysis facilities themselves saying we often don't see people for the first three days after they're discharged. How can I be held accountable for the readmission when I didn't even know they were in the hospital yet? You know, that kind of feedback is the kind of thing we'd like you to begin thinking about how to incorporate.

If a small workgroup of you want to help us think this through, we'd be delighted. Because again, it's brand new. It's just a way of kind of almost giving you insight into where I

know we're going so we can collectively think about how to make this work.

Did you have a comment, Kyle?

MEMBER CAMPBELL: Yes, I'm just trying to think of this in the context of the example of our current readmission measures. So I think we would -- we start with an academic medical center and maybe like a very small sample that might get at this, but later in the process we'll be doing a national dry run, which I think speaks exactly to what you're trying to get at here. So is there a pathway or has there been thought about a pathway like we kind of have to get you the measure when it's on the timeline for the project, but maybe within a year we'll have sort of all this information to get at NQF+. And so, is that something that you guys are thinking about in terms of additional information?

DR. BURSTIN: Yes, and actually one of these we are thinking -- we didn't really speak much to this because we're mainly focusing in on this particular issue today -- is that the

Endorsement+ will be something you could come back for over time. So you could -- a measure gets additional testing. Measure's out there. It's being vetted. You now move up that criterion, those criteria and then you can be -- and then you could get Endorsement+. So it doesn't have to be something just at the initial submission. Potentially off-cycle, or maintenance cycle, or whatever it may be.

And part of this; actually it was interesting, was Beth McGlynn, who's on our panel and the chair of the MAP, who actually made the point that one of her hopes was that if Endorsement+; again, we'd love a better name -- if Endorsement+ is something that will be viewed as valuable in terms of people selecting measures for particular programs and that's something you would shoot for, then will that be again a potential sort of driving force to have maybe maintenance be a bit more meaningful, right?

If there's an opportunity to move to that upper level, then you could work with those

who fund you and others to say we really want to shoot for that because that would really help us moving forward with end users and the specific programs in which they might be used. But great suggestion.

Please?

MEMBER KEENAN: So just as a follow-up to that, there are some measures where there's not data currently available. And so, it's not possible to really calculate the measure and distribute performance results before the measure is kind of out there in use. So do you mind speaking to that a little bit?

DR. BURSTIN: I mean, our preference is you give them performance results and data, but I think if nothing else you could give them the second part there, which is at least feedback on the measure itself as an opportunity to get their thinking even if you can't yet give them data.

MEMBER POPOVICH: Yes, so during the committee meeting who came with Endorsement+ --

and I like the name right. So I endorse Endorsement+.

MEMBER POPOVICH: What percentage of measures did the committee believe could reach the status of Endorsement+ right now? And then what's the scaling factor over the next two to four years of what would be expected?

DR. BURSTIN: It's a great question and we probably need to go back at some point and do that analysis. We think very few measures right now, Matt, would meet this bar, particularly because of this one. I think we have a fair number of measures that on the other ones around testing, reliability, validity and evidence -- a handful of those, maybe 5 or 10 would fit that. But some of those haven't been vetted. So because it's the new addition of the vetting, I think over time we'll see that increase.

MEMBER POPOVICH: So this criteria is one that they thought was probably going to be the higher bar?

DR. BURSTIN: They thought it would be the new bar to meet, the challenge. Although I will tell you; and we've had lots of discussions as part of our measure developers meetings and others, getting to actually doing formal testing of validity is not easy either, particularly for outcomes. And we've heard you there.

So we'd also love to collectively think about what are other ways some of that validity testing could be done in ways that may not be quite as onerous, but at least get the -- I mean, many of you have heard it: committees just don't like face validity. They don't kind of buy it. How is that a valid representation of quality? Because a group of folks said it was?

So we accept it. We think it's a reasonable starting point, but just the reality is I think as quality measurement has kind of gotten ratchetted up as being more and more important in the broader scheme of things, I think it's not surprising you would see more of a forcing function on wanting to see the measures,

particularly those that might get used for those somewhat higher stakes uses, people selecting doctors, people -- hospitals getting paid, et cetera.

But I think there might be a need to kind of ratchet that up a bit. But we recognize this won't be easy, which is why we're not making it a must-pass. It's kind of more of a signal to you to work with us, help us think through how to do that.

So, an opening.

MS. JOHNSON: Other thoughts? Question from the --

MR. TILLY: Right, Amy Bennett asks why are measures endorsed with an exception to evidence not eligible?

DR. BURSTIN: Yes, I think the thought is that we want to have measures that are -- truly meet the highest bar across the board. And an exception for evidence would not, in our esteem, be measures that would be at that level.

I think one of the challenges has been

that this whole evidence question is something we still need to wrap our heads around. At times particularly for some of the more cross-cutting measures, less clinical measures it is harder we know to get the evidence. And so those are the ones I think at times very likely to get exceptions. I think an early example in palliative care was spiritual counseling for palliative care. I mean, what kind of evidence are you going to need to get to that? And yet a fair amount of evidence that it's a good thing.

But I think what we don't want to do is have measures move forward for which the clinical evidence in particular isn't there, or it's equivocal. I think that's the hardest time. And the thing we often collectively get the greatest heat for is when clinicians and others are told to do something when the evidence is in play, when one week we read estrogen is good; next week estrogen is bad; next week estrogen is good and bad. That's probably not the right time for a performance measure, so we would think

those would be ones that don't quite reach that threshold.

Anything to add?

MS. JOHNSON: Outcomes measures deliver --

DR. BURSTIN: Right. Good point. Karen mentions outcome measures don't have that problem because we don't require quality, quantity and consistency evidence for outcomes, just the rationale.

MEMBER KAYE: This is Toni again from the PCPI. Thinking on that evidence exception question, as a lot of the gaps in programs there's really a push for things like patient engagement and care coordination, and those happen to be the things where you don't generally get that grade A randomized trial evidence. So do you worry that -- I think it's likely that a lot of the Endorsement+ measures could wind up being those clinical process of care or outcomes as they come, so it might be harder if programs are simultaneously encouraged to use maybe

Endorsement+ measures as meeting the highest bar, but very few measures that maybe meet those gap areas. I think they're kind of disadvantaged from the start to meet that criteria, so I think it could make the gaps worse.

DR. BURSTIN: I think it's a fair comment and it's something we'll keep an eye on. Again, we want to emphasize as well endorsement is great, right? I mean, that's a great bar. We're not saying there's anything wrong with that bar, but I think the question is particularly for some programmatic uses at times might there be any -- and I think we'll see it, right, if we really begin to see that there's a need for some of those really important gap area measures in the higher stakes programs.

Will that -- again, we're open to it. It's just our thinking at point A fully recognizing a lot will change I think in all of our worlds around measurement and measure development in the next couple years. But great questions. We hear you.

All right. I think we're done. Back to you.

MS. ISIJOLA: Okay. So we're going to go to a 15-minute break and we'll be back in 15 minutes. Thank you.

(Whereupon, the above-entitled matter went off the record at 10:35 a.m. and resumed at 11:00 a.m.)

MS. ISIJOLA: All right. We're going to go ahead and get started. So just a few housekeeping items.

Again, the slides are actually available on the webinar if you are dialed in or logged in via the webinar, but we will also send them out following this workshop, including the recording for your use.

Also, on our agenda it does say networking lunch. We're going to do a working lunch so we can kind of get you guys out of here earlier. I think that will be helpful for everyone.

But now we're just going to turn it

over to Karen Johnson and she's going to talk about assessing validity.

MS. JOHNSON: Thank you, Wunmi.

I am very excited about talking about validity, but I'm one of the geeks of the group, so we'll see if we take our whole hour or not. And we don't have to if you don't want to. But I hope you'll play with me. I have a few discussion questions throughout and I'm really kind of interested in hearing what you're thinking about validity.

So, I'm going to start out by reminding everybody where validity fits in our criteria. It's the second sub-criterion under scientific acceptability. And you notice that I like to use color. So when I think about our criteria, I really do think of three separate pieces to validity. One has to do with evidence. Are your specs consistent with evidence? One has to do with testing. That's the second one there in orange. And then the other ones we grouped together and we call those threats to validity.

So recently somebody asked me -- I've never heard of this, threats to validity. And I don't know that we actually use that term very much externally, but I'm using it today. So if you hear us talk about threats to validity, this is what we're talking about. Exclusions. We're talking about risk adjustment, meaningful differences, missing data, comparability.

Okay. So again, just a reminder. I know you guys know this and if I gave you a quiz, you would be able to tell me all of those threats to validity, right?

All right. I'm also showing you exactly what we state in terms of what we're looking for in our criterion guidance. Some people I think, and probably rightly so, find it hard to find stuff from NQF. So I just wanted to tell you that in our 2015 criterion guidance document, pages 12 to 14, you will find what we're looking for in terms of validity. And again, I've used color to pull out a couple things that I think is really important. Data

elements, measure score, face validity. So other important things in there, too, but those are things that we're going to kind of focus on today.

Any questions about this? You guys know what I'm talking about when I say the 2015 evaluation and guidance document? I see some yeses. Any nos?

Okay. Good. All right. So testing. And today's talk is going to be mostly about testing, even though all that good stuff about threats to validity is in there, too.

So testing key points. You guys already know this. Validity we think of as referring to the correctness of measurement. And we are interested -- and we allow two different kinds of validity testing.

Thank you, Reva.

Reva is keeping me honest up here. One is empirical analysis, and that can be done at either the data element level or the score level. Score, I tell our new people that come in

that's the results of the measure, because sometimes that word is even a little strange, right? NQF-speak.

It's important that the testing be done for the measure as specified. So that's one of our things that we actually do require. It has to be as specified. Face validity and the measure score is also accepted.

Okay. Any questions about this?

All right. This is old hat. All right. So let's talk about data element validity. Usually it uses patient-level data, right? So we're looking at the actual data from your patients or the data that's being aggregated up. We would like to see testing for all of the critical data elements, not just one overall kind of result. At minimum, however, we would like to see it for numerator, denominator and exclusions. So possibly three groups.

And when there's data element validity going on, we actually expect to see high agreement, right, because we're looking for

accuracy. Okay? So again, that's what we're looking for in terms of results. If based on an instrument or scale, we're looking for validity of the instrument or scale. So that would count as data element validity for those kinds of measures.

And you guys probably know this and probably have taken advantage of it, if you've demonstrated data element validity, we don't require additional data element reliability. So you can get two birds with one stone on that one. Okay?

Now this is the fun part. What we often get is some percent agreement and nothing else. Who knows what the problem with percent agreement is?

(No audible response.)

MS. JOHNSON: Elisa and I could be talking about something that I absolutely know zero about, but we could agree on it quite often if it's just a yes or no, right?

Sam?

MEMBER SIMON: Basically it could be thrown off by low frequency.

MS. JOHNSON: Prevalence has a lot to do with it, too, right. Right.

So then we say how about kappa scores, kappa statistics? What's that do for us? You guys didn't know there was going to be a test, did you, a quiz?

Okay. That helps us with that chance agreement. Even if I don't know anything about anything, I could still agree probably half the time, right, if it's a yes or no option. Right?

Okay. So kappa statistics is better than just percent agreement by itself, but probably still not exactly what we're looking for, although we usually let it slide by. What we'd really like to start seeing is sensitivity specificity, negative and positive predictive values. Now I'm not going to quiz you on that because I don't remember the details of how you calculate all these things.

Yes, it all comes from the kind of two

by two tables, right? So if you can get the kappa statistics you can get the sensitivity specificity, right? Okay. So this is what we would like to see more of. So maybe you guys can be helping us out on these.

The other things that we often get is very little explanation of the method. We computed some kappa statistics and here you are. And I'm not really being that facetious. Sometimes it is that brief. But having things like how big of a group did you look at? What were you pulling your data from? What methods did you use? And are you considering XYZ the gold standard? Otherwise, we have to assume, and sometimes we assume incorrectly, right?

But if you tell me that you had an expert nurse doing some abstraction, is that the gold standard? Why don't you just say that it's the gold standard? Then everybody understands and it's very transparent about what you did and what you're expecting. Right?

We often, as I mentioned before, see

one value with no explanation. So you might have six critical data elements, but I see one kappa statistic. That's not uncommon. But I don't know what the kappa statistic is for, and neither will the committee. But you guys will know, right? This is your stuff. Right? But it's like when you write something, you can't tell if you've like made a typo or you've left out a word. Somebody else has to look at it sometimes. But that's what happens. Sometimes we get this stuff and we don't quite understand.

Finally, no interpretation. We have to put this into some kind of perspective for people, especially people who aren't stats people on our committees. Is a kappa of 0.5 good? Is a kappa of 0.75 good? Right?

All right. So any questions or comments about this? You guys may be -- you're looking a little bored. I don't want to belabor this. The only thing down at the very bottom, and I almost forgot it, is values aren't great. At minimum we'd like you to speculate why or why

not. If you don't have good values, it tells me that it's not valid, right? So do we want to endorse a measure that doesn't look valid?

And then sometimes you might speculate. Do you go back and retest? Again, I do understand that things cost money and it might be hard to do, but it is a question that comes up. You've demonstrated that something isn't working the way that you had hoped. You fixed it. Did you fix it? Right?

So any comments on this? Does any of this surprise you? Kind of what we're getting -- I know none of you guys are the ones that are giving me one value with no explanation, right?

Is there anything that we can do to make it more clear what we're looking for?

Ah, Karen. Yes?

MEMBER DORSEY: So probably everybody here tries to their best due diligence to provide all the information that you all need, but I wonder what importance emphasis you all put on the length of these. I feel like over time the

length of these submissions grows. We give more specific and more detailed information for the committees to consider. I mean, there's always been some guidance about length, but I'm just -- how are you all thinking about that?

MS. JOHNSON: I'll tell you what my interpretation is, and it might be interesting to get Reva's or Sarah's or somebody else's. I think it should be as long as it needs to be to be clear about what you did, and no longer. Yes, says the people that read them. You know, a lot of times I don't think necessarily it's the testing pieces that are too long. Often we get really, really long evidence pieces and it's because it's that old -- when you're a freshman in college and you write a paper and you're not quite sure what you're supposed to say, so you say everything that you know and you hope that that is it. Sometimes that happens. And that really lengthens the submission.

So when we say with evidence -- and I'm kind of getting off track here. When we say

with evidence we want a summary, we really mean a summary. We're expecting a paragraph or two. We're not expecting pages. We're not expecting pages of citations, that sort of thing.

So I don't know if that helps or not, but brevity -- we still have -- I don't know how many people are using it, the What Good Looks Like document for the testing. And almost everything in there is short, short, short. There's hardly anything long in those docs.

So, Sam, I just saw you get a microphone. You got a question?

MEMBER SIMON: Just a quick one. Does the committee ever ask to see a power analysis? I mean, you could do an abstraction on 20 cases and get great agreement, but it may not be sufficiently powered.

MS. JOHNSON: Right. I think I've seen it or heard it once or twice. In general they don't ask for that. We have toyed with the idea is that something we want to ask you guys for and make that kind of bring it out more. And

so far we've decided not to because we do understand that your power analysis might tell you need 300 and you can only realistically get 50. So we don't require it because at this point it might be a little hard. This is for data element testing.

For score level testing I think none of that matters, right? I mean, you're going to have -- if you're going to do score level testing, you've got a lot of data is our assumption. You agree or -- Sam said he was going to agree with everything, but I kind of don't --

MEMBER SIMON: Well, I think maybe what could be done is to say how much power you have given the sample you have.

MS. JOHNSON: Ah, yes.

MEMBER SIMON: So that might be --

MS. JOHNSON: Yes.

MEMBER SIMON: -- something else to consider.

MEMBER SIMON: That kind of stuff I

think is very helpful. What we usually have on our committees, to be honest, is some people who haven't ever taken a stats class and don't really get it, but there's going to be some people who may as well have a Ph.D. in stats, because they're going to be looking at everything. And we all have our strengths and weaknesses and things that kind of jump out. So to the extent that you can satisfy that one person who's going to be looking at every detail, I think it's a good idea. Yes, right now we're not requiring it.

MEMBER SIMON: Okay. Great. Thanks.

MS. JOHNSON: Yes. Any other questions? Yes?

MEMBER GEPPERT: Just wondering about your experience of people submitting sensitivity and specificity data.

MS. JOHNSON: Yes.

MEMBER GEPPERT: And, I mean, is it sensitivity I think that's particularly difficult to estimate sometimes trying to identify --

MS. JOHNSON: Yes, I've actually never actually done a sensitivity analysis, I'll be honest with you. So I don't really know how hard it is. I know sometimes what we have seen is people pulling from the literature. You see it more I think -- I've seen it mostly the claims kinds of data.

PARTICIPANT: Compared to the chart.

MS. JOHNSON: Yes, claims compared to the chart, and there are lots of papers published on that sort of thing. And, yes. Yes. So I guess I'm not quite sure why it's particularly hard. Prevalence has something to do with it and --

MEMBER GEPPERT: Well, I think particularly for a claims based measure, for example, if you identify a case in a claim, then you can go and pull the chart and determine whether that's a false positive or not. But it's hard to find things that you're not --

MS. JOHNSON: Yes.

MEMBER GEPPERT: -- identifying in

claims.

MS. JOHNSON: Well, you get this thing -- you know, when we said earlier that Reva and I are constantly emailing back and forth and some of us, I'll be doing one of the PAs and I'll say they didn't do all the critical data elements. And Reva will bring me down and she'll say they probably used claims to find their denominator. And then I'm like, okay, I'd still like to see it for the denominator, so I'd really like you to do a random sample of your medical records and go that way. And Reva's like get real, Karen, that's not going to happen.

So, yes, you are correct. When you're doing it that way, when you're using claims or some other -- a registry or something to do data case finding, yes, it can be --

MEMBER GEPPERT: And you approach that scene with some sort of stratified sample where you try to identify a smaller group that is very -- has a high likelihood of being --

MS. JOHNSON: Yes.

MEMBER GEPPERT: -- a false negative and then --

MS. JOHNSON: Yes, so that you can make sure.

MEMBER GEPPERT: -- and then maybe a larger part of the sample that has a very low probability.

MS. JOHNSON: Yes. I mean, it is harder. I'm looking at a couple of measures right now where they didn't do case findings, so my question to the committee is are you -- do you feel pretty confident that that can be reliably extracted kind of in -- without that empirical data. They haven't done that kind of pulling of records that you're describing.

So, yes, but you're right. Reva has -- we have harder graders here sometimes and easier graders, and we really are trying very hard to be consistent among staff. So I've gotten a little easier on my grading and testing, probably a little harder on me grading for evidence. So that's how it works.

So I have kind of like skipped myself. Any surprises as to what I've mentioned so far just in terms of our criteria, what we're looking for, what we've found? How many of you usually do data element validity, or are you guys just now just skipping to score level? What do you think about data element validity? Is it valid? Do we need to care? What other stats or methods might work? I've mentioned a few. Do you buy the kappa? We kind of let kappa go through. Our notes would say kappa is not good enough. We really want sensitivity specificity, that sort of thing. Any -- yes, Karen?

MEMBER DORSEY: Since we mostly work with claims, we don't really need to deal with data element validity except for hybrid measures that come forward. And it seems to me that it's still useful to demonstrate data element validity for EHR data elements.

I mean, I think that whether or not you can use sensitivity specificity depends on whether you can identify some kind of gold

standard or conceptually set up the question that way, which is not always a natural fit for EHR data elements. The approach that we've taken is to say that the most important thing is that data values match, but that's at the data element level, not at the -- like an eCQM score level.

MS. JOHNSON: Right.

MEMBER DORSEY: So I mean, it's hard for me to wrap me my head around a statistic if I can't put together a conceptual framework for how there's a gold standard to compare to. I agree kappa is -- doesn't feel very satisfying.

MS. JOHNSON: Yes. So I'm a little curious, and I want to pick at it just a little bit, do you feel like the EHR data are not the gold standard? I mean --

MEMBER DORSEY: Sorry. So I'm talking about when we're validating extraction, right, saying that, yes, you can extract through electronic query accurate data elements as compared to the front of the chart --

MS. JOHNSON: Okay.

MEMBER DORSEY: -- where people are actually entering the data. So that's the validity I meant when I was talking about --

MS. JOHNSON: Okay.

MEMBER DORSEY: -- EHR data element validity. And in that instance I guess the front of the chart is technically the gold standard, but it's -- I mean, you're kind of stretching the intent, right, of sort of the conceptual test of sensitivity specificity to apply it. It feels kind of arbitrary and unnecessary when all you really care about is that the data elements match --

MS. JOHNSON: Right.

MEMBER DORSEY: -- and that you have done it in an adequate sample.

MS. JOHNSON: Right. You can tell I'm not a clinician because I've never actually had a chart either.

MEMBER DORSEY: Right.

MS. JOHNSON: I know that there's a difference between the front and the back of the

chart, so apologies there.

Sam?

MEMBER SIMON: I think -- so for the EHR measures that we're testing we are -- I think this brings up the point that it would be really helpful if measure developers explain, contextualize what do their validity results mean? Are they taking results from the front of the chart? Is this sort of a broader view across -- like we're comparing extracted EHR elements from everything that's in the EHR --

MS. JOHNSON: Right.

MEMBER SIMON: -- so whether something's written down in a notes field, the idea being that our validity results will tell you if you rely on just structured elements, are you getting the whole story or not?

MS. JOHNSON: Right.

MEMBER SIMON: So having that -- some narrative. I don't want to make the forms longer than they are, but it does kind of speak to why some narrative is probably the right thing to ask

for --

MS. JOHNSON: Right.

MEMBER SIMON: -- about how to interpret your results.

To answer your question about kappa, I totally understand why you would want -- why kappa alone is insufficient. I mean, I think you want -- the whole idea is you want the whole two by two table, right? So that's what the sensitivity and specificity --

MS. JOHNSON: Right. Right.

MEMBER SIMON: -- results give you. So I think it makes sense to ask for those.

MS. JOHNSON: Yes.

MEMBER SIMON: And really for eMeasures the thing we run up against is that doing -- we would love to be able to do measure level scores.

MS. JOHNSON: Right.

MEMBER SIMON: And we just don't always have -- sometimes there's not a great thing to test, but how do you measure for measure

of care coordination? We can usually give you the data element validity; less so with a great outcome to really benchmark those scores against.

MS. JOHNSON: Right. And you're stealing my thunder. That's the next set of slides, or almost the next set.

Yes, let's go on. We can't stop until we talk about face validity, right, and get into the juicy stuff.

So face validity, judgment of whether on the face it the results appear to reflect quality of care. Right? So again, we're remembering that we want our results to be -- allow for accurate conclusions. So that's where accuracy correctness comes in.

It is subjective. It is the weakest form of validity. And why? Because one set of experts may not agree with the next set of experts. Right? And Reva's saying that's what happens when you get on the committee because you've got a different set. You may not agree that something is valid.

So what are our requirements? We do allow it. We have toyed with getting rid of it. So far that hasn't happened. We'll see if it ever happens. We don't know. But right now we are saying that it is a -- we want to see a systematic assessment of the score from the measure as specified. So this is -- we want to know if the results are an accurate reflection of performance and whether the results can distinguish good from poor performance.

So what we sometimes get; this is fun, we get people to tell us whether their experts thought the measure was a good idea. And again, I'm not being facetious on some of this stuff. We get a lot of face validity about the construction of the measure. Did you use the right ICD codes? Do you think you can collect that denominator? That sort of thing.

But remember, we're looking for face validity in the measure score, not the construction of the measure. Right? We get feedback on the feasibility of the measure. It's

not bad information. It's just not what we're looking for for validity. No description of the experts involved.

Sharon, you're smiling.

Little to no description of what was assessed. A general statement of results. Everybody thought it was great. I'm being a little facetious on that one. Or sometimes no results at all. We did face validity, period.

Yes. Yes. Going back to your original question about how long should the form be, I probably have a little bit different flavor maybe than Reva, but I really like the idea of transparency and I really want people to be able to go back to the forms even a year or so later and understand what happened. And it's kind of the same thing Reva was saying, but, yes, if you know the literature, you may know that this panel thinks this is great and they really did a great job, but somebody going back later that doesn't know that literature and doesn't have that paper by Dr. So-And-So in front of them won't know

that. So that's why we like to see the actual results.

And I don't know if I -- yes. So what we'd like to see. Details about what was asked and how the assessment was done. A list of the experts involved with their credentials. And then actual results of the assessment. So here's an example. Nine out of the ten experts agree that the measure reflects -- measure results differentiate good from poor quality. None disagreed, or whatever you got to. Right?

So any surprises on this? How many of you guys do face validity only? Not any -- and it's not -- don't be embarrassed. It is a valid form of validity and it is something that NQF accepts. How do you decide whether to do face whether empirical? Is it money? Is it sample size? Data? You flipped a coin that day and this is what you ended up with?

MEMBER GEPPERT: I do see a little inconsistency with -- the vetting is a good idea and an important new component to validity, and

yet somehow a structured face validity process is no longer useful. I mean, I'm not seeing a huge difference between those two processes --

MS. JOHNSON: Yes.

MEMBER GEPPERT: -- and at least the face validity process, if it follows something like a delta, or something like that --

MS. JOHNSON: Right.

MEMBER GEPPERT: -- it's very structured, it's multistakeholder --

MS. JOHNSON: Right.

MEMBER GEPPERT: -- you're not dealing with biases. I mean, it could potentially be more useful than simply getting some unsolicited comments from people that are -- could be negatively impacted by a measure.

MS. JOHNSON: Right. Right. It's a good point. Yes, I mean, we spent a lot of time, to tell you the truth, and I don't know how I forgot to tell you; Helen took care of it, but we really were thinking that the vetting thing would go under validity because we kind of saw it as a

face validity, if you will, of the people being measured and others, which I keep forgetting to add on, but -- the "and others" part.

But again, you saw that we put it in usability and use instead. There was actually an extra piece of it that -- of the definition that we ended up not doing because that was kind of pushing way too far. So I don't know if that quite gets --

Helen, do you have anything to add?

(No audible response.)

MS. JOHNSON: Any other comments about face validity? Yes.

DR. WINKLER: Just in response to you, Jeff, the face validity here and what we tend to see tends to be people that are very much involved and proximal to the measure development process. So the expert panel, the advisory panel, the whoever. Now sometimes they'll go to -- and I've seen face validity assessed at say a subcommittee of a specialty society that weren't necessarily developing the measures, but they go

to their -- to another committee and have them do it. Great. Super.

But again, it tends to be folks that tends -- that have that frame of reference as opposed to the vetting, which is of those being measured. And I think that's the critical difference. I mean, your expert panel advisors may ultimately end up being measured, of course, but nonetheless I think the target and the -- for the concept is what's the critical difference between the two.

MS. JOHNSON: Thanks, Reva.

All right. So empirical testing and the measure score. So this is where I'll give you some of our philosophy and you guys can push back, if you want to, or maybe you'll agree.

I would say that to some extent our thinking has evolved. And, Kyle, I think it was you that talked about different types of validity. We have that whole little list of criterion, predictive, concurrent, all those things that you learn and then you forget until

you have to apply it or something like that.

Now when I look at things, I don't really care what label you put on it, if you call it predictive or whatever. I think of it really all as construct personally. And the difference is what are you comparing it to? If you're comparing it to some kind of gold standard, then that becomes what people sometimes call criterion validity. Right? So that's how we're thinking about it. Different ways to validate, but all assessing some kind of hypothesized relationship or relationships on the measure results to the results of another measure or measures based on the knowledge of the underlying constructs.

So don't get too bent out of shape on the word "hypothesize." We're not necessarily saying it has to be hypothesis testing in the stats sense of the word. We see it as a theoretical exercise. And I think many of you have already talked about this. Sometimes what we hear is I can't really talk to you about validity because we have to get our statisticians

on the phone.

And we would say the statisticians are there to help you crunch the numbers and maybe come up with your methodology, but you guys as the developers, the subject matter experts, should be the ones who are coming up with those relationships and deciding what you should be looking at, or maybe what you shouldn't be looking at, right, and what you should expect to find.

So here's how we see it: Link a concept of interest to some other concept by hypothesis or a construct. Usually there are many and we might -- I might get some pushback on that, because I've often heard that there's not any, right, or not any that you can actually do. But we'll talk about that.

The hypothesis should indicate the direction of the relationship. Do you think things are tracking together? Do you think they're tracking differently? Tell us, right? And it's theory. You should have some idea about

the strength, the expected strength. And we'll do some examples so this makes a little bit more sense, but the idea there is do you really -- if you're looking at a gold standard, you probably expect a pretty strong relationship, right? That's what you're hoping for. You want to validate, right? But if you think what you're looking at is maybe related but maybe a little weaker, well, you should be able to articulate that.

So you link it in your mind. You do something with it with some numbers and some data, some kind of method. And then you look at what you found. You could do formal statistical testing and look at significance. You can look at your effect size. You can look at directions. This works with correlations, those kinds of things as well. Right?

If the expected relationship is found, then you at least have some comfort in thinking that your hypothesis may have been correct and that your measure is valid. Okay? If you don't

find what you expected, it could be a problem with your measure or possibly your hypothesis, or both. Right? But validity is really -- there's not really one -- there's usually not going to be one major analysis that's going to totally validate a measure. Validity is -- I think of it as like a brick in a wall, right? You just keep kind of adding to it so that you feel more and more comfortable over more and more tests that you have a valid measure.

Now I also realize that in the real world you may only have one brick, but at least theoretically you could continue to add things on. You're doing measure instrument development. You might look at it in one population versus another population with English and Spanish, etcetera, etcetera, and you'd validate in all of those areas. Right?

All right. So that's our thinking. I've already hit this. Different methods could be used. Correlate. Look at differences and means, progressions, whatever. And what you pick

depends on what kind of data you have and what your level is that you're testing type of data. So I kind of merged in a little bit of data element testing here, but -- and of course your type of data.

Let me stop there and see if this at least seems reasonable. Maybe this is not how you've thought about it yourself. Maybe you have other words that we should take instead of these.

You want to do an example first and then we'll come back to it? Let's do an example. Okay.

Very recent example. This week, as a matter of fact, we had a couple of these come through. They did some really nice score level testing. So the hypothesis was -- so the measure was about NICU admission temperature of low birth weight babies. But the hypothesis was that hospitals that had more babies with low temps when they're admitted to the NICU will have a higher NICU mortality rate. Okay?

So they've come up with this

relationship. They think there's something going on between -- and the literature supports it, which is always good. Right? So they have a relationship there. They have a direction. They say it should be higher. Right? It's a positive relationship. They don't really talk too much about the strength size that they're expecting. I guess they could have gotten even more specific.

But here's the results. Three hospitals. They actually found a dose response relationship here. The lower the temp, the higher the mortality.

And, Reva, do you want to -- this was your measure.

DR. WINKLER: Yes, this was a -- it was very elegant. Small number of hospitals, of course, but essentially what it was is the larger the number of babies, the higher the mortality. Duh. Therefore, as a proxy for predictive mortality, neonatal temperature on admissions to the NICU is looking to be a very valid assessment

of the care provided. And so it was a very elegant, very simple, not a lot of data, but told the story quite nicely.

MS. JOHNSON: If they had messed up somehow or another; it's hard to imagine with this measure because it's a pretty simple measure, but if they've messed up the construction of the measure or something, some kind of goofball thing happened, and the results didn't show that, you would -- what would you think? Would you change your hypothesis or would you think, boy, something's wrong with maybe how I constructed my measure perhaps? Or maybe I have to go look at a couple other hospitals because there's something weird with these hospitals and luck of the draw got me three weirdos.

Okay. Is this making sense now as to what we're talking about? Let's do one more.

Unexpected complications in term newborns. We predict that our measure will be highly correlated with a neonatal admission to

the NICU. This was a different measure.

So what they did is they showed by week, which is actually very nice, the -- I think these are odds ratios of expected -- unexpected complications.

Is that what this was? Do you remember?

(No audible response.)

MS. JOHNSON: I should have put that on my table. Yes, it looks like them.

And so, unexpected complications. Admissions to newborns. And what you see here is a dose response, but really a U-shaped curve. Woo hoo. And there you are. There's your U-shaped curve. It's doing what they expect it to do. They have validated their measure using a measure score. This one in particular I think; Reva, remind me if I'm wrong, they didn't have this data in house, right? Did they -- is this the one that they borrowed data or had somebody else run some data for them?

DR. WINKLER: (Off microphone.)

MS. JOHNSON: No, this is -- okay.

DR. WINKLER: (Off microphone.)

MS. JOHNSON: Okay. All right. Apologies. I was thinking there was one that they actually used somebody else's data to help them validate.

DR. WINKLER: The interesting thing about this is this U-shaped curve has been applied to just about every outcome in obstetrics when you relate it to gestational age. So that's what makes this particularly an elegant result.

MS. JOHNSON: So we're now co-presenting.

(Laughter.)

MS. JOHNSON: All right. Interpretation. So they gave us some tables with some numbers in them. They gave us a pretty chart. We can therefore conclude that severe and moderate -- I'm not going to read this out. So they gave the interpretation and then they went a little further and they said this demonstrates that our metric successfully captures and

quantifies neonatal morbidity in term newborns. So they really laid it out. They really interpreted it for us.

They also note that what they found matches the LIT. And they did another test of correlation. Both sets of -- this is referring back to the table. Both sets of odds ratios, there we go, demonstrated high positive correlation after further reinforcing results and conclusions. So basically the had one hypothesis, but they did two separate analyses to look to see if what they were expecting actually happened.

All right. So just for a second; and I realize I'm preaching to the choir here, but let's go back. Here we go. Linking a concept to some other concept by hypothesis. That's really like highfalutin sounding, isn't it? But really they just said we think this should track with this other thing over here. They did some testing and then they looked at the results and made some conclusions.

All right. What do you think? I think I'm at my what-do-we-think stage.

Sorry. This thing is really slow.

Oh, I was going to tell you some of this. Let's come back to that in a minute.

Any surprises thus far? We'll go back and I'll tell you what we sometimes get.

Thinking of this as assessing relationships, does that help when we -- when people ask us what do we mean by score level validity? Does that make sense? Is that language working? Is that how you think of it?

Kyle?

MEMBER CAMPBELL: I think it makes a lot of sense the way you have it laid out. And you asked sort of like when do we consider face validity? And I think it sort is the last result. If we don't have an opportunity where we think that we have something that we can compare against and there are those -- particularly the outcome measures, there are those issues where there isn't something suitable as a comparator to

do this type of work, that's kind of when we would rely on face validity. And we probably would do face validity anyway, because we're interested to see what the panel thinks.

MS. JOHNSON: Yes.

MEMBER CAMPBELL: I think you have it -- I think you've captured it really well. And I think the important thing is the conceptual relationship, right?

MS. JOHNSON: Right.

MEMBER CAMPBELL: So if you had a strong correlation but conceptually it didn't make any sense, you could put it down, but what would it mean? So I think understanding that, what is that relationship telling you as you go in? Your hypothesis is really important.

MS. JOHNSON: Right.

MEMBER CAMPBELL: Yes.

MS. JOHNSON: And you got to my -- the last bullet there, what are some of the barriers. So we have heard that it's hard sometimes to find other measures to compare to. So I would say one

thing, an outcome doesn't have to compare to another outcome. You can link it back to process measures, if you have them. Right?

But are there other barriers? Is that the big one? Is that insurmountable? Does it depend?

Oh, we have a question. Yes?

MR. TILLY: So, yes, perhaps related to barriers. Purna asks how do you do empirical testing when you don't have the right data elements?

MS. JOHNSON: Okay. When you don't have the right data elements? Yes, so I think what she means is -- she doesn't have the right data to do score level testing, is probably what she means. And I think that is the barrier, that people are saying they don't have the data. If you think that you're admissions are related to your discharge planning and you don't have that information, you can't do that analysis. I think that is the problem. And that is where Kyle goes back and -- you always have face validity to fall

back on. Yes.

How many of you actually are running into this problem? You would love to do this kind of stuff, but you don't have the data?

(No audible response.)

MS. JOHNSON: Okay. Sam, is that in eMeasure kind of world?

MEMBER SIMON: (No audible response.)

MS. JOHNSON: Okay. So you don't have data anyway, right?

(Laughter.)

MS. JOHNSON: Sorry. I couldn't resist.

Now it surprises me a little bit sometimes when people have claims data or registry data and they can't do it, or they -- but again, you have to have the right relationships, the ones that make sense. So I can certainly see it.

Anybody else want to -- have you thought about getting data from other people if you don't have it?

(No audible response.)

MS. JOHNSON: So I don't think I've answered this question, unfortunately. That's why you guys are here to help us answer.

Reva, do you have a --

DR. WINKLER: (No audible response.)

MS. JOHNSON: No answer? If you don't have the data, you don't have the data. I would say don't necessarily say it has -- it doesn't have to be criterion validity. Right? It doesn't have to be outcome and outcome. It could be -- like one that we see every now and again is -- the fancy term I think is -- what do you call it, the group, differentiating group's validity?

So if you know that you've got certified recognized stroke centers versus others that aren't, you would probably expect your measure to be higher in your recognized centers than your not recognized centers. That could be one. That might take a little bit of outside data to know which one is which.

Hey, you guys must be getting tired.

Hungry? Okay. All right.

Let's go back and I'll tell you a little bit of what we often get. Let's see, did I miss any? No. Okay. What we often get. We get a naming of a method. Usually the naming of the method is correct. Not always, especially when it comes to things like predictive validity, because different people mean different things when they say predictive validity. Right? They might actually mean concurrent validity, but I'm using some kind of regression model. So I'm predicting something. Right? Which is fine.

A correlation table but no explanation and little interpretation. So here's my -- I copied it from SAS. Here is it. Yes. What we'd like to see. Description of the hypothesized relationships. Which measures did you choose and why? Don't assume that everybody's an expert and would know your thinking. Let us know why you thought that that was the case. Let us know what you thought the case was. If you expected it to be positive, negative. You expect to see a high

correlation or a strong association. Or maybe not so much. But you should know that a priori, right? You guys are the experts, so you would know this. Right?

And then interpretation of findings. How are -- does or doesn't this validate your measure? Little bit tricky if it doesn't. But even if it didn't, you may say on reflection this may not have been the strongest hypothesis or, yes, the expert panel -- yes. Often go to the expert panel. So these are some of the things that we would like to see to really round this out.

Otherwise, what happens is we have to assume. We're writing these preliminary analyses, so we're making assumptions about what you're thinking. Or we're chasing you down. It doesn't always work well when we make assumptions because we're wrong sometimes. Or the committee is confused, what have you.

All right. Is there anything left of this? Everybody's hungry.

Extent of validation. I've already mentioned this. Often demonstration of validity should come from a series of studies. And this is just kind of typical. If there isn't a gold standard, then you would expect that you would need to do this more and more to really build your case of validity. Again, I understand there may not be money or whatever to do this, but in a very kind of theoretical kind of world this is what we would do. Right?

More studies strengthen the evidence of validity and it's built over time. This one's a little tricky. Even one study with unexpected results with respect to a gold standard may invalidate the measure. But that's what you want to know, right? You guys don't want to put forward and invalid measure, because what we want to do is drive improvement. We don't want to -- so this is why you test.

And then NQF criteria also require consideration of potential threats to validity. So we are going to talk about threats to validity

very quickly. I've already mentioned those. They're listed in our criteria. Patients inappropriate excluded. Differences in patient mix. Scores that are generated with multiple data elements or sources or methods, and then systematic error. That can be unintentional or intentional. Lots of different ways that validity can be threatened.

We ask you to assess potential threats. That's what we're asking for. Mitigate them if you can. And some people do, right? You might impute if you have a lot of missing data, for example.

So what would we like to see? So for exclusions at minimum we'd like to at least know how many you're excluding. It would be even better if we knew the variability across measured entities. Now if you're excluding hardly any, it kind of doesn't matter, right? But if you're excluding a pretty good portion, then is it uniform across entities or not?

Preferability sensitivity analysis of

results with and without exclusions. That's not always possible. Sometimes you don't even have that data if they're excluded in the first place. But sometimes people can do it and they will say it doesn't matter, or it mattered a little bit, but not too much, or whatever.

Risk adjustment for outcome resource use. Some process measures. As Lynn and PQA folks told us yesterday, there are adherence measures we initially considered at process measure, but they've realized that risk adjustment is probably appropriate there.

We want to see empirical analysis to demonstrate that risk adjustment isn't needed, so it is possible to put forward an outcome measure that doesn't have risk adjustment. We have to convince people that you don't need to risk adjust, right? So we want to see data.

And then if you are planning on risk adjustment, particularly in light of our SDS trial, we would like to see conceptual and an empirical approach. So the conceptual piece is

the rationale for the SDS factors. And we want to see that even if at the end of the day you decide that you're not going to put them in there anyway. We want to know why you considered them.

And then discrimination calibration statistics. Is your modeling -- it's usually modeling. We don't see very much risk stratification. It's usually statistical models of some sort. Is it doing what you want and does it fit the data well?

Meaningful differences. At minimum we'd like to see variation amongst the measured entities. Even better would be some kind of statistical analysis, not just saying that they're different, but they're really different in some meaningful way, either clinical or statistical, or both. Right? Sometimes we get things that are statistically different, but not meaningful in some kind of clinical way.

Comparability if multiple data sources or methods are used. I'll admit that this is one that we hardly ever see anything on. It makes me

nervous if a measure is specified at claims and some other way, because how do I know that it's fair to compare this one and this one? And we would like to see that.

Missing data. How frequent are they? How are they handled? Demonstration that results are not biased.

Any surprises here on what we're looking for?

(No audible response.)

MS. JOHNSON: Often we'll get NAs and people will say, yes, I had a bunch of exclusions, but I'm not going to tell you anything about them.

Okay. We've got a couple minutes left. If you could change anything we do regarding validity, other than not asking you about it in the first place --

(Laughter.)

MS. JOHNSON: -- what would it be? Any ideas? It could range from you're doing it perfectly, good job to get rid of face validity.

So, Kyle, have we got you or what? Any ideas?

Oh, we have a question. Yes?

MR. TILLY: Yes, Purna again asks is risk adjustment still needed if a measure is just to be used for reporting purposes and not for accountability?

MS. JOHNSON: If it is an outcome measure or a resource use measure or even some process measures that you feel like that patient characteristics could bias the results, then yes.

Oh, let's hold on a minute. I haven't got anything from our audience about what you would do if you were NQF and you could change anything you wanted. There's got to be at least one idea.

Jeff, yes?

MEMBER GEPPERT: Throw out an idea. So for both reliability and validity we sort of -- we treat them like thresholds, right?

MS. JOHNSON: Yes.

MEMBER GEPPERT: You're either reliable or you're no.

MS. JOHNSON: Yes.

MEMBER GEPPERT: You're valid or you're not. And that's probably not true.

MS. JOHNSON: No.

MEMBER GEPPERT: They're more like -- it's aggregation.

MS. JOHNSON: Yes.

MEMBER GEPPERT: So I think that introduces some unpredictability of the -- into the process where one committee might consider something reliable and valid and another wouldn't. And so I think anything we could do to help standardize those type of assessments would add to the process.

MS. JOHNSON: We've actually thought a lot -- you probably have noticed, I think other than our EHR testing where we say it has to be more than one, I think that's the only threshold that we put on anything. And the reason is that nobody can agree on the thresholds. So if we said we have to have a 0.7 of reliability, there's going to be people saying we'll never get

to that, but it's still a reliable measure. And that's a rule of thumb that has not basis in reality. So it's been really hard. You're right, that would really help.

What we're trying to do and what we're hoping -- it would be interesting to see if you guys think it's working, is the preliminary analysis that we're doing. We're trying to be more consistent internally. We've already admitted that we're -- some of us are different graders and we're trying to be more consistent here. We're hoping that that's making our committees more consistent in what they're thinking about and how they're ultimately voting. It would be interesting to hear your feedback on those things.

But if you have ideas about how we could come up with thresholds, that sort of thing, I mean, we're willing to think about it. The last time -- we tried about two years ago and it didn't fly. They didn't like that.

MEMBER GEPPERT: Actually I'll go the

opposite direction.

MS. JOHNSON: Oh, okay.

MEMBER GEPPERT: I would embrace the continuity of the concepts of reliability and validity.

MS. JOHNSON: Yes. Okay.

MEMBER GEPPERT: And then I would try to incorporate that into my sort of decision making process.

MS. JOHNSON: I think it's -- that actually works a lot better when people interpret their results, but often all we get is our reliability estimate was 0.63. Our c-statistic was whatever it was. So you might be right.

I think the other thing that is interesting, we really do think of reliability and validity as context-specific. So we talk about it. And we're kind of lazy in our language. We talk about is a measure reliable or is it valid? And really it might be reliable with this piece of data at this point in time and not so much over here. We understand that.

Right? We sort of understand that. And we kind of let that go, because we're never going to be able to test every possible permutation of everything. So that's why some of us think NQF stuff is the low bar.

But any other ideas?

(No audible response.)

MS. JOHNSON: Intriguing idea, Jeff. We'll think about it. Like how can we do that?

DR. BURSTIN: One option might be to just pass out cards so you don't have to raise your hand and say what you want us to change. I mean, you could do it during lunch and we'll collect them. So you could -- feel free to say whatever you like. I mean, about validity would be nice, but more broadly would be nice as well.

MS. JOHNSON: Sure. Yes. Yes, we'd love to know your feedback. It would be interesting for us to know if you guys think we're on track with what we're asking for. We do realize that you guys think -- and it is burdensome to fill out our submissions, our

forms, etcetera. We're trying to make that easier. We feel like for the most part that we're asking you things that are important to ask. So if you don't think so -- when we have people thinking about their reliability and validity testing particularly, usually what we say is think about the data that were used and how kind of -- what's the scope of that data? Think about the method. Was it appropriate. Think about the results. Does it seem reasonable? And that's kind of our guidance. It's not rocket science really. I don't know.

Anybody else have any final thoughts? We'll give you cards. Looks like lunch is coming out.

Thank you guys for talking back to me. It made it a little more fun for me. We would look to you to give us any feedback you have. If you have it, we'll consider it. Thanks.

Oh, that's right. I forgot to pass.

(Laughter.)

MS. ISIJOLA: Okay. So we're going to

break for lunch and spend about 15 minutes. You can grab your lunch and then come back and then we'll go into our next session. Thank you.

(Whereupon, the above-entitled matter went off the record at 11:56 a.m. and resumed at 12:19 p.m.)

MS. ISIJOLA: Okay. Thanks everyone for holding on. So just some housekeeping, some more housekeeping. Jean-Luc actually just placed some cards at your tables. I know there was some discussion of providing some feedback, so that's an opportunity for you. If you don't want to state it publicly, use the cards and provide that feedback to us and we'll use that and integrate it as we continue to improve our work.

Also, following this workshop we'll be sending out a survey, and that will help us really continue that effort. Your participation in this endeavor is really crucial to how we do our work, so we'll be sending a survey. What worked for you during these past two days, what hasn't, what are some of the things you would

like more information? I know Amy Miller mentioned some of the discussions.

And sorry to call on your, but some of the discussion is meaningful to some of your colleagues.

One of the things we'll be doing is putting some of these recordings on our pages and have it as an on-demand. So it may not be useful for you per se, but it may be useful for your colleagues who work with us, really thinking about the testing or the validity. So stay tuned for more there.

And with that being said, I'm going to turn it over to Reva. I know over the past two days we've talked about CDP and MAP integration, and this is an opportunity to talk through that. Reva will speak as well as Sarah Sampsel. And we also have one of our CMS colleagues who will be able to join in on that discussion as well.

Reva?

DR. WINKLER: Well, it wouldn't be Washington without the alphabet soup. And so,

CDP, the Consensus Development Process. I don't know familiar you all might be with the term that we use very readily internally, but that's our endorsement process. CDP has three letters. Endorsement has many more. So CDP it is. And then MAP, the other process. We work together on both of them. Let's just talk about what those two entities are.

They are two distinct multistakeholder processes here at NQF. So the consensus development process, or CDP, is our formal process to evaluate and endorse measures. This is what we have done since the beginning of NQF back, oh, 2001 or so. It's focused on NQF's measure evaluation criteria. And the timing of the CDP projects is not as predictable. It has to do with the various scheduling and the contracting for funding. And so, we do issue the periodic calls for measures to bring them into the CDP process. So that's -- and it is based on using the NQF membership and other members of the public in a very multi-stakeholder fashion to

contribute to building consensus around what measures should be endorsed.

The other entity at NQF, other large entity at NQF that is another multi-stakeholder process is the Measure Applications Partnership, or the MAP. And the MAP began about five-and-a-half years ago to provide annual input to HHS on the selections of measures for use in federal programs. All right? So where the other is around endorsement, this one is around selection.

And the measures under consideration by HHS for possible use in federal programs later on is a very tightly scheduled activity. The MUC list is released every year by December 1st and the Measure Applications Partnership, or the MAP, provides feedback no later than February 1st. So you can count on those. It's an over-the-holidays, over-the-new-year cram and jam. So, but it's very predictable because MAP season has replaced holiday season for most of us.

(Laughter.)

DR. WINKLER: So, and it's the

focusing on the use of measures in specific federal programs. And so, you're aware that during the course of evaluation for endorsement while we talk about use and usability, we really aren't talking about use in specific programs. Very different from the MAP, which is looking at specific programs. So these two processes were, especially when the MAP first started, relatively distinct here at NQF. I mean, they actually lived at the other side of the building and the staff were distinct. But it didn't take too long for us to figure out how much we needed each other to understand the activities of one and the information that needed to flow back and forth.

And so over time, and particularly over the last two years, we have really made concerted efforts to integrate the communication flow between these processes. And so, you will be involved as developers in some of this information flow. And I'd like to try and just describe it to you.

And so, I've got various versions of

this, because it's a circle and it's an ongoing, multi-directional simultaneous communication information flow thing. Okay? It doesn't fit very well on a two-dimensional surface. But if we just start at one point, if you will -- and that's with measures that have been through the consensus process and they're endorsed by NQF. Okay. Well, sometimes those measures end up on the MUC list. Yay. And so what we are making a more -- much more concerted effort is to provide the information that the steering committee, standing committee used to evaluate that measure to the MAP. And how that the MAP has seen it, oh, my God, their appetite for it. They're voracious. Okay? It's like I want all that stuff.

So one of the things that we really -- it's now standard process for us for any of the measures that are endorsed by NQF that go to the MAP is that all the information from the endorsement process: all of the comments, all of the evaluation, all the voting, all the

everything that happened, they have access to all of that. So, and we believe that that's a particularly useful input to the MAP.

Now, the MAP has a lot of conversation and a lot of opinions about use of those measures in very specific programs. And as part of the MAP discussion there's a lot of feedback. There's a lot of, gee, did the committee think about that, or, gee, it would be really nice if the committee would consider this, or, gee, we really think you didn't mention this and it's a problem because of X, Y, and Z.

And so, that feedback from the MAP will go back into the measure endorsement and maintenance process so that it tends to be a circular kind of flow of information. It's possible that the MAP feedback is such that we need to do an ad hoc review of the measure. We would obviously notify you that that's where we were going with it, but sometimes the concerns around the use of certain measures in certain programs have prompted sufficient questions that

we need to kind of stop and take another look.

So because we have standing committees, that becomes an -- we now have sort of the vehicle by which we could do that ad hoc review. If not, if it's just good information, good feedback, it becomes one of the most important feedback loops that we currently have to understand what's going on around the use of the measure. And that information gets fed back to the committee at the time of the next maintenance review.

So the flow of information between the two is intended to be continuous. And this happens all the time because the CDP processes are not tightly scheduled. They're happening pretty much all the time. So we're constantly pulling the MAP input in. So this is happening on a regular basis. This is not even close to being linear.

Okay. So what are other options? Well, we know that on the MUC list when it comes out every year there are a lot of measures that

have never been through NQF. Okay. So they come into the MAP and the MAP is sitting there going, hmm, okay. And one of the struggles they have is they don't have the information that we -- the depth of information and the details that we have. And they really -- they like that. They like it a lot actually.

But what happens is if the MAP is looking at a fully developed measure, they may say, okay, we support it, but conditional on it going through NQF so that it can have a more vigorous look at the measure and the measure characteristics. And so, what we are doing new, if you will, is to reach out to the developers of those measures and said look, the MAP said, hey, this would be a great measure, gave it a conditional support. We are going to try and bring this into NQF at the next available opportunity. And so -- and make sure that you're aware that that is the situation and we may or may not know what our upcoming projects are, depending on where we are in a contracting cycle,

but to let you know that you're on our radar.

Our funding proposals are also now aware of these measures so that we can talk with CMS and some of our other funders about saying, hey, there are these measures out here that were proposed and the MAP liked them and we really need to bring them in through the endorsement process. And so, that can help drive or structure some of our upcoming projects. So that's a relatively new thing we're using to try and help establish the upcoming project cycle and topic areas. And then whatever feedback the MAP have had about a measure is fed back into the committee when it goes through the CDP for potential endorsement.

So as you see, between MAP and CDP there is an important flow of information. Both are informing the other. And the thing that's made this work particularly well over the last year or two is the fact that we no longer have distinct staff. We all work on both. And so, we don't have to -- we know where the information on

one is -- could be found for the other. It's not a -- they aren't separate entities in house anymore. They are all just part of our Quality Management -- our Quality Measure Department activities.

And so, I did want to talk a little bit about some of the specifics very briefly, because you as developers might get pulled into some of this, and you certainly would want to be aware of what might be happening with your measure as it's flowing around NQF. And so, realize that the endorsement evaluation will go to the MAP, and the MAP may provide feedback on endorsed measures.

So it's very possible, as I had to do after the clinician MAP this year, is go to a couple developers and say, yo, the MAP is -- raised concerns or issues or questions around X, Y and Z. Your measure is up for maintenance this year and we will specifically ask the committee to address those things raised by the MAP. And this is something we're doing right now. It's

currently happening.

If it were to be something really, really big and the measure was not up for a soon evaluation, it might prompt an ad hoc review. Of course we would tell you that and you would be part of that process. So realize that we really -- we have to really see what the situation is with the measure and with the issue, but either of those things is likely going to happen, and we have several examples ongoing right now.

So when you look at the measures on the MUC list -- because I'm sure you're all aware CMS sort of has an open call for measures that's currently open right now and through I guess July 15th. And so, HHS creates the Measures Under Consideration list; we so love it, the MUC list, and that is what goes to the MAP.

The MAP may, as I said, recommend the conditional support. And if that's the case, as I said, we would want to do the outreach to you all and let you know that that's what's happening.

Now, one of the comments we had yesterday was interesting, because what we're seeing, particularly on the clinician side, maybe a little less so in hospitals, but certainly some on the PAC/LTCs -- we're seeing a lot of measures come to the MAP that are not fully developed. They're still in process. The amount of information we get that's on the MUC list about that is so minimal we really don't know what that means. We just know they haven't checked the box saying fully developed and tested. It's maybe there's alpha testing, maybe there's beta testing, maybe this and that.

I know I tried to reach out to the developers last year and I would hear, yes, the measure is being tested in the upcoming year in your registry. Okay. Great. But it wasn't done yet. So one of the things that the MAP overall and the Coordinating Committee struggles with is the fact that some measures just aren't ready to be implemented in programs yet. We need a little bit more time to finish the development. So

we're struggling with some of that limited information that's available about some of these measures, because frankly for some of them there's not a lot more than the title description, numerator, denominator and we're done. So it is a little bit difficult.

But we will be wanting to interact with the developers of those measures, and particularly the developers of the measures where the MAP feels warm and fuzzy about the measure, at least in concept, but if it isn't fully developed yet, where are you in the development process? And then the added nuance, which I probably need to update, is the thought of, well, if you're still really early on in testing and perhaps need some assistance with testing, and it's a really great measure, particularly if it would make a really good eMeasure or something like that, perhaps we can connect you up with the incubator. So we do want to integrate all of these processes within NQF to take advantage regardless of where we are in these -- whether

it's MAP or CDP or the incubator or whatever else we're happening to do.

I want to now turn it over to my friend Sarah who's going to tell you a little bit about how this integration with the CDP, a specific -- one of the standing committee and the PAC/LTC MAP Workgroup function over some of the measures in that particular topic area.

MS. SAMPSEL: Thank you. And so, I'm going to make a couple of statements and kind of tell the story of what's been going on over the past couple of years between person and family centered care and the PAC/LTC. But then our colleagues from CMS are on the phone, Mary Pratt and I believe Tara McMullen and perhaps some others who represent the actual programs, so work across both committees as well.

To cut to the chase, the biggest issue is communication. And so, it's not only those of us at NQF understanding the measures and the programs on the PAC/LTC side or on the CMS side, but then understanding kind of the timeline and

how they're coming through our CDP process as well.

So really a shining example is the implementation of the IMPACT Act. And for those of you who are not familiar with this, which I was not familiar with up to a couple years ago, this was a bill passed in 2014 and signed on October 6th, 2014. And what it requires is standardized patient assessment data that allows data element uniformity, quality care and improved outcomes, quality -- comparison of quality care across post-acute settings, improved discharge planning, exchangeability of data and coordinated care. That should be really easy, right?

And so, what we were finding and have been experiencing is as CMS has been going on their timelines and implementing the quality measure domains and the respective timelines for implementation into the programs these measures were starting to come through the endorsement process as well. And so, while the PAC/LTC MAP

was educated on the IMPACT Act and understood why the considerations for the measures and why, as Reva just referred to, there might be measure concepts, there may be measures earlier in development, they were really just CMS alerting these measures are coming forward.

But then on the endorsement side -- so on the CDP side these all came through Phase 2 of person and family-centered care, which was focused on functional status. All of a sudden we had 28 functional outcomes, functional status measures being brought to a CDP project.

And so, our committee members were -- some of them were certainly well aware of the IMPACT Act, but it was bringing those pieces together and working with our colleagues at CMS to say, hey, can you come and talk to us about the IMPACT Act and would you just make sure that when we're in deliberations -- because one of the questions then comes up during the CDP process, especially on usability of use, how are these measures going to be used? As these QI measures?

Are these reporting measures? Are these pay-for-performance measures? Those are really things that we need the developer to answer and should not be NQF.

So a couple things happened. One the CMS representation along with their developers at the panel meetings, but in addition to that during the previous MAP process and now current MAP process; I'm a senior director, or had called in to be able to listen to and provide that direct feedback from the MAP. And that's a process that Reva just explained is always improving, always evolving.

But as we're getting ready for our next phase of person and family-centered care work, while these measures aren't coming through, it's a great opportunity to keep the standing committee informed of what's going on, what they may see in the future, what happened in the past, kind of what's the steering, what are those gaps even identified by the PAC/LTC, the duals, and some of those workgroups that filter over both?

So really kind of what happened. And this is the simplified example. So quality measures were introduced on the MUC list for PAC/LTC programs. The programs that we look at for post-acute and long-term care are SNF, in-patient rehab, home health, long-term care. We also look at hospice, but that's not covered under the IMPACT Act. The vast majority, if not all, were encourage continued development.

Some of them were adaptations of endorsed measures, which we talked about yesterday in the variation conversation, but really we want to see more setting-specific testing before we make these recommendations for approval in the program. Thus, those should come back at some point to the endorsement review and the ability for person and family-centered care. But that again is that feedback loop. Now we need to update when they come back, this is what the MAP said about these measures.

Overall what we think the benefits are: first of all, understanding the intent of

the measures. So when you're filling out your measure information form, when anybody's filling out their measure information form, what is the current use of the measure, what is the planned use of this measure? This is one of the areas we can pick that up. We know that CMS is planning to use this in a program or has already put it in rule or whatever. But knowing that really helps our committee understand what's going on and how the measure will be used.

It also helps us -- as I already mentioned, what's coming down the pike? So by following what's happening on the MAP, looking forward to the next year, next two years, again our negotiations, our discussions with CMS. Well, then we can plan projects around this. We know what's coming up. We know we're going to get at some point in the next couple years another 20 measures or whatever that is.

And that's not a hard number in any way, but just saying that we have an idea of what's coming forward not only for perhaps the

most relevant committee, but maybe there are some other ones that are filtering to other committees as well. And that way we can start preparing internally on education to make sure that other folks don't have the same experience of kind of catching up behind the scenes.

And then as already mentioned, the use and usability, implementation information, overall intent rationale for the measure is huge. And that's where we really -- it does help to learn kind of what is CMS doing? What are they working on? Where are they going with these measures? Because it ends up impacting the overall endorsement process. And we had some issues with appeal -- measures being appealed for some of these issues here, which we then identified as maybe these folks don't really understand how this all works together and how we are sharing information.

So on my part I wanted to go ahead and pause, but wanted to make sure that the mics are open for Mary and Tara and others at CMS who

might have additional thoughts and comments on the process.

MEMBER PRATT: This is Mary. Thank you. This was very thoughtful conversation and information that you're presenting. And Tara McMullen and I are here.

I guess, well, one potential aspect to point out with the IMPACT Act is that it's not always the case, but in this case we had a clear mandate for what we were building and what we were supporting in that development of the measures. So we were able to have a source where we could turn back and look to time after time to say are we meeting the intent, or are we straying from it? And that gave us a real solid foundation.

As huge as the IMPACT Act is in terms of what it encompasses for the post-acute care settings and with delivery of quality care to beneficiaries and others, we at least have a starting point. And we found that orienting the MAP and as many stakeholders as possible to what

we were doing and why kept the engagement of people involved in the process and helped us along the way.

And now Tara McMullen is here and she and Alan Levitt attended a workgroup with our contractor and she has similar points to make, if there's time.

MS. SAMPSEL: Sure. Please go ahead.

MEMBER McMULLEN: All right. Hi, Sarah. It's Tara. And thanks to Mary. And I think Sarah and Mary kind of teamed on it.

The IMPACT Act is kind of -- it's put us in a new direction where we do have a solid foundation, or like kind of we have a flight path for the work that we are to complete within a certain and specified amount of time. But at the same time it has put CMS and NQF with the MAP in a position where there has been an increase in measures to be reviewed, an increase of work, and, many times something that we heard from the last MAP, an increase in understanding in where CMS is going with the measures and why these

measures?

Overall the 2015 MAP, Alan Levitt and I took away that it was a really great meeting. WE thought that it was the best MAP meeting that we've been to, and we've been to every single MAP. The dynamics of the panel were really nice. We felt that -- a complaint that we've had in the past had been time management, and we felt that it was under control. And we felt that this was the most prepared CMS has been and that there was a really solid interaction between CMS and the MAP panel members. And we really actually enjoyed ourselves. And we hope that the 2016 MAP is the same way.

And there were a couple things that we wanted to bring up in the light of the IMPACT Act under the guise that there is an increase in work and the increasing amount of measures that we will submit to the MAP probably won't subside. We will probably be increasing many, many measures in the years to come. And these measures not only are measures developed under

the IMPACT Act, but they're measures developed for quality reporting programs.

And that kind of drives home -- the first point I want to make is that the focus for CMS for the Measure Applications Partnership is that we're submitting measures to the MUC list to be reviewed by the MAP to get some feedback on the measures that we're proposing to use potentially for our quality reporting programs. So at least Alan and I have not seen this as an endorsement mechanism.

We hear the word "endorsement" a lot. This is our pre-rulemaking process. And we are attending the MAP to gain insight into how the panel members feel the process is or how our measures would actually enhance our quality reporting program. And that's one thing that we took away this year was that we felt as if we were having to walk through the measures, which isn't an issue, but walk through the measures as we were facing an endorsement panel, the panel that we actually submit our measures to for the

consideration of endorsement.

And we were wondering if we could have some discussion about that, really about the intent of the MAP. And if the intent is now to -- if it is kind of more like an endorsement or are we viewing our measure submissions as submissions to enhance our quality reporting programs, because that was something that we -- we were under the guise that that was the intent.

And we also had a question about knowing if the IMPACT Act has really increased the work load and the work flow for everyone from CMS to NQF, if there was an opportunity for CMS and NQF to coordinate activities to decrease the burden of the MAP members. The one thing that we heard from a lot of the MAP members last year was that they were really appreciative that CMS was open and transparent and that we were there in person. And they were very appreciative of that, but they said there is so much material here, I don't even know what I'm looking at. There is so much going on.

And Alan and I brought this up with Michelle Geppi, but we thought maybe it would be a great idea prior to the actual MAP meeting to have a meeting, like a -- via telephone or whatnot with the MAP members to walk them through our measures so they'd know what the measures are when we walk into the MAP and they're aware of the concepts. And they don't have to tell us what they think about the measures, but that they just have an understanding of what the measures are. We feel like that might save some time.

MEMBER PRATT: And we think that as measure developers preparing their measures in such a way that helps facilitate that kind of understanding and discussion with the members of the MAP and NQF that that is the kind of predominant work that will help facilitate everything maybe that flows from that point.

MEMBER McMULLEN: Yes, exactly. Thanks, Mary. And then just to kind of wrap it up, we, in bringing a point that Mary brought in and this whole point about how can we coordinate

with NQF to be more efficient, the one thing that has occurred in the last couple years was that since we're on such a rigorous timeline in our pre-rulemaking cycle, kind of going by when JIRA is open, when JIRA is closed, when the MUC list is open, when we can provide our measures, it's proven to be pretty difficult for DCPAC, the Division of Chronic and Post-Acute Care.

And as we all know, in 2014 we had exercised an ad hoc Measure Applications Partnership to review the first kind of wave of the IMPACT Act measures. And this year we were updating our MUC list up until the last moment of the MAP. And in consequence of that, for the Hospice Quality Reporting Program, but really for some of our measures such as Medicare spending through beneficiaries the MAP didn't have the most accurate and detailed measure specifications and that just kind of created some confusion.

And it's not a big deal because we were able to speak and walk through them. But I think it just opens up an opportunity for us to

talk through coordination and collaboration and understanding that NQF is on a rigorous timeline. So is CMS. We kind of feel like we have our backs up against the wall. And we're working on measures up until the last moment. So the life cycle, well, it keeps going. It doesn't stop. So just because a measure goes on the MUC list doesn't mean we stop enhancing it and polishing it, testing it.

And last but not least, we just wanted to bring up the idea or maybe some discussion around the public comment period at the Measure Applications Partnership that occurred in 2015. We really appreciate public comment. It's actually one of my favorite things in the whole process because I think it really highlights for us some of the directions that we should be doing as an agency in our measure development.

The public comment period was a bit different this last year. It was just a bit different. I think the words, a lobbyist effort, was thrown out by multiple parties. And so I was

wondering, Sarah, if you guys at NQF had a discussion about that and if that discussion would be appropriate for today?

And so, I'll leave it to Mary to close, but those are some of the main thoughts that CMS is having. But we appreciated the process last year.

MEMBER PRATT: So, hopefully this is sort of a bird's eye view of some of the unique aspects that we encountered on this very successful year with the MAP and NQF. Just some new challenges and really having a measure development team that we work closely with that is agile and knowledgeable and can work together with CMS as well as NQF to meet the needs. That's all we have.

MS. SAMPSEL: Well, thanks, Mary and Tara. We appreciate that feedback, and obviously there are some take-aways that kind of NQF and CMS can take back, but certainly exhibits the point that there is a lot of collaboration that goes on between CDP and MAP and understanding

each other's work and work flows, and they do impact each other on extremely tight timelines. But we're doing our best to do that, and that's something that is obviously extended to any measure developer, as Reva mentioned, with the outreach on letting you know that programs are coming up, et cetera. But it's always a back and forth between organizations to get that ironed out.

So with that -- and just, Tara and Mary, I'm not sure if you're staying on, but we have a group here at the NQF offices, all measure developers, and then a number folks on the phone, also all measure developers. So we had some questions that we're again looking for your feedback on. And these questions are -- and feel free to notify us online if you have questions.

But some questions for developers include how closely do you follow the annual MAP pre-rulemaking activities? I would add on top of that have you always been aware there are two distinct NQF activities between endorsement and

MAPs? Have any of your experienced this integration of CDP and MAP and how is it affecting you? Or how could you perceive it to be affecting you perhaps in the future?

And then you kind of heard some of the issues and examples of things that happened between NQF and CMS, between CDP and MAP. So do you have any ideas for best strategies to include developers and the flow of information?

As Tara mentioned, there are a lot of times that measures are undergoing constant work. And in this case I think with the hospice measures the MAP met on a Monday and Tuesday and we had received updated specs the Thursday or Friday before. And those timelines might be off a little bit. But again, that's an exchange of information. So what strategies do you have to kind of keep us updated so that we can update our panels in house?

So we'll open that both to folks in the room. And then if there's anybody online that has questions, Jean-Luc's monitoring as

well.

MEMBER DOMI: Hi, this is Marsida, AHCA. Thank you very much for the overview with regards to the MAP-CDP integration. It has been really informative and wonderful to hear this type of work.

I just had a really quick question. I'm not sure if you know, with regards to the JIRA the form encompasses numerator, denominator and so on and so forth. It's not as extensive as an actual NQF application. However, I was wondering are there any talks with regards to -- since we're in the theme of integration, are there any talks with regards to getting the two systems to integrate together so that you, I don't know, can click something and it automatically populates something? I know I'm wishful thinking, but just a question.

DR. WINKLER: I'll wish along with you. At this point when you're talking about the hardware infrastructure not -- I mean, we talk about it, but the likelihood of anything in the

near terms doesn't seem real high.

But one thing that does happen is we have met and talked with CMS during -- multiple times over the last couple of years in terms of the information that is included, the fields that they use in JIRA. And so, what would be particularly more useful for the MAP, the kinds of information they're looking for?

Our biggest struggle actually when we get the MUC list, which is generated from the data submitted through JIRA, is so much of the data is not there. I mean, there are very few fields that are required and there are a whole bunch of fields that we'd sure like to see. And what the MAP gets is a bunch of blank fields. So there's very little information.

So I realize that it -- because they're not required, they're not required. But realize that the more information, the better it helps the measure, because then there just isn't a lot of unknowns where people fill in the blank in the way that suits them. The more you can

provide that information that's requested, it's really much more helpful for the MAP to go beyond the simple required fields.

But we do have conversations back and forth with the folks that are managing the JIRA as well as in terms of trying to keep them somewhat aligned so that at least the way we ask the same question can be similar. And I think we can continue to have those conversations, but at this point -- not next week.

Questions from anybody else?

MEMBER DORSEY: So I had two questions. One was that we're -- maybe this is more of a comment, but we're always racing against the clock because there's such a long period of lag between the conclusion of measure development and success of implementation. And so, we're often really pushing to get things on the MUC and before the MAP so that the clock to implementation can start ticking. And because of that I think -- and we also sort of -- we like to provide a complete specification report for the

MAP. And because of that I think we are often pushing right up to the limit of the MAP meeting in terms of getting information.

And so, I mean, I'm saying that to say I think we are interested in how to streamline that and create efficiencies given the time constraints and are open to conversations about how best to present information to the MAP and what information is critical and what might be more than they need. So I wanted to say that.

And then the other is just I'm curious about how you all are thinking about how -- sort of formally including information about the MAP recommendations into the committee's deliberations. Like what kind of conversation or addition to the conversation are you anticipating from that? I'm trying to sort of get my head around how that goes.

DR. WINKLER: Right. I mean, we're still just seeing the first couple as we've started really doing this this spring. It's under use and usability. We've kind of created a

category of feedback loops. And it won't be exclusively the MAP. It's just right now that's our best source. We're looking to try and create whatever feedback loops we possibly can to get information from the field on how it's going, because that's really the fundamental question under use and usability. Is it being used? How's it going? What are the problems or -- and successes?

And so, like I say, right now your most ready source of feedback that we can easily get to is the MAP. And so we're including it in that. And I think that it's just one more bit of information. And I think it becomes a lot more -- it probably has a lot more import for measures that have been around awhile, maintenance measures, because it's like what's the experience with the measure? It's been recommended and it is in use in this program. And there's just more to talk about to raise that, where before we were just silent. The MAP didn't exist and nobody had any feedback and we just didn't go there.

And so, I think it adds something. And I think we're still evolving what we're able to say. And I think that the more information the MAP has to work with, the more they're able to say. But the limited amount of information on the measures at this point that often comes through JIRA onto the MUC list is so limited, they start speculating all over the board.

MEMBER DORSEY: Right.

DR. WINKLER: And it's hard to translate that into something concrete that you really want a committee to pay attention to.

So again, the more information there is -- and I know last year -- it's a very, very short time frame from when the MUC list becomes available and before we take to committee and -- or the workgroup. And, but I know last year I tried to contact many of the measure developers just -- you know, is there anything more? Is there anything more you want to tell us? Because I know that the submission deadline for that information was back in July, and things change

in four or five months. And so, if there's some kind of update, we're happy to collect that information and to be able to provide it to the MAP. It's just we're running on such a short time frame.

So perhaps just be aware. If you know you've got a measure that could be coming up on the MUC list, we're probably going to try and get in touch with you and see if we can find out if there's anything new in your world. Particularly if you're saying the measure is being tested, it's like, well, perhaps you finished in the intervening time frame. We don't know that. So again, it's all about information flowing as readily back and forth.

MEMBER DORSEY: So, because my brain always goes to the worst case scenario, I'm just thinking about -- there may be an instance in which a measure is recommended for use in a program and the MAP declines to support that recommended use. And so, I would think that in those cases it would be important to give

instructions to committees about how they're supposed to think about that in terms of their -- whether to make a decision to endorse.

DR. WINKLER: I would agree with you because it's -- that's what I'm saying. When you have very superficial conversations with minimal information, it's really hard to understand that and then be able to transfer that information. So the more we're able to have data-driven, information-driven conversations that can readily transfer, I feel much more comfortable saying, well, the MAP said this. Well, to tell the committee the MAP didn't like it. Why? I don't know, they just didn't like it. I mean, what do you do with that?

So they were concerned that the denominator didn't include this population, blah, blah, blah. I mean, the details become -- makes it much more valuable information exchange.

Anything else from anybody? Kyle?

MEMBER CAMPBELL: I just wanted to follow up on what you were saying, because we

follow the MAP very closely. So I think that that issue of the lack of information fostering dialogue that isn't necessarily valid in relation to the measure is an important point. And I don't know what your thoughts are about fostering more dialogue with the measure developers per se to -- I know there's limited amount of time, and that's another thing. Like I feel like the MAP gets into areas where the CDP really should be focused.

And so, what are the different criteria sort of between the MAP and the CDP? And if they're evaluating things with the absence of information, it's really hard to make that --

DR. WINKLER: Yes, the Coordinating Committee and the MAP will be the first ones to agree with you on that.

MEMBER CAMPBELL: Yes.

DR. WINKLER: Which is why they have a preference for measures that have been through NQF simply for the information that's available. And so, we struggle with it. Like I say, I think

it's reasonable for us, given the limitations on time, that if it's at all possible to get -- to interact with the developers before we take it to the workgroup to do that.

Again, as I say, it's always over Thanksgiving. We usually get the MUC sent to us on Wednesday afternoon before Thanksgiving. Again, as I said, those of us at NQF, it's not holiday says, it's MAP season. And so, it's that time frame thing that just makes it really hard.

So, yes, I'm more than happy to send you an email, but I'm not at all surprised if you don't answer it over Thanksgiving. So that becomes some of the limitations of just using that particular time frame. So I think we're happy to try to do it, but realizing we're both limited on each side of the fence.

Yes, I think that's the thing everybody feels a little -- I wish I had a little more to work with on the MAP side.

Anything else from anybody?

(No audible response.)

DR. WINKLER: Okay. Thanks. Anybody on the phone? Operator? Yo-ho.

OPERATOR: Okay. To ask a question, please press star then the number one.

(Pause.)

OPERATOR: And there are no questions at this time.

DR. WINKLER: Thank you. Ms. Wunmi?

MS. ISIJOLA: Okay. And thank you to your CMS colleagues for just joining us. I think the input that they provided was helpful for you as well, but it also helps us to think about ways that we can more closely align our work with the CDP and MAP work.

So with that being said, I think this is the end of the rodeo. And also hopefully you filled out your cards. We'll definitely take that into consideration, as always. We have plenty of opportunities for you to engage with us, whether it's via email from our maintenance inbox, measuremaintenance@qualityforum.org, whether it's contacting project staff on specific

project areas, but also we have our submitting standards page, which has all of our resources. If you have any question about anything regarding our measure maintenance work, we encourage you to utilize that.

As we mentioned over the past two days, there's a lot of new revisions and upgrades to some of our policies and processes. We'll definitely be informing you of that during our various channels on our web site, through emails. Also following this meeting, again we'll definitely send out the recording as well as the slides.

But I'll stop one more time to see if there are any questions about anything that we've discussed, any final thoughts and comments, what you felt about this workshop over the past two days. Was this helpful? Informational? I see nods. Anyone on the phone, any comments?

Okay. Well, thank you again everyone for participating and sticking with us for the past two days. Happy Cinco de Mayo and enjoy

your weekend.

Meeting's adjourned. Thanks.

(Whereupon, the above-entitled matter went off the record at 1:06 p.m.)

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Measure Developer Workshop 2016

Before: NQF

Date: 05-05-16

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R Gross

Court Reporter