The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 9:00 a.m., Dennis Andrulis, Co-Chair, presiding.

PRESENT:
DENNIS ANDRULIS, PhD, MPH, Co-Chair
MARSHALL CHIN, MD, MPH, FACP, University of Chicago
LOURDES CUELLAR, MS, RPh, FASHP, TIRR-Memorial Herrmann
COLETTE EDWARDS, MD, MBA, CIGNA HealthCare
LEONARD EPSTEIN, MSW, Health Resources and Services Administration
KEVIN FISCELLA, MD, MPH, University of Rochester School of Medicine
DAWN FITZGERALD, MBA, Qsource
ROMANA HASNAIN-WYNIA, PhD, Northwestern University Feinberg School of Medicine
ELIZABETH JACOBS, MD, MAPP, University of Wisconsin, Department of Medicine
JERRY JOHNSON, MD, University of Pennsylvania School of Medicine
FRANCIS LU, MD, University of California, Davis
MARY MARYLAND, PhD, MSN, BC, APN, Chicago State University
ERNEST MOY, MD, MPH, Agency for Healthcare Research and Quality
PRESENT(Cont'd):
SEAN O'BRIEN, PhD, Duke University Medical Center
NORMAN OTSUWA, MSc, MD, FRCSC, FAAP, FACS, New York University Hospital for Joint Diseases
GRACE TING, MHA, CHIE, WellPoint
DONNA WASHINGTON, MD, MPH, VA Greater Los Angeles Healthcare System
ELLEN WU, MPH, California Pan-Ethnic Health Network
MARA YOUDELMAN, JD, LLM, National Health Law Program

NQF STAFF:
HELEN BURSTIN, MD, MPH, Senior Vice President, Performance Measures
ROBYN NISHIMI, PhD, Consultant
TAROON AMIN
ALEXIS FORMAN
ADEELA KHAN
NICOLE McELVEEN
ELISA MUNTHALI
C-O-N-T-E-N-T-S

Welcome and Agenda 4
Topics from Prior Days Discussion 6
Discussion on Health Literacy Measures 13
Measure Gaps 26
Risk Adjustments 69
Disparity Sensitive Measures Assessment 118
Next Steps 178
Adjourn
CO-CHAIR ANDRULIS: Okay, we've gone from a day of being led by a great task master to now you have to deal with an Austin slacker. So I wish you the best of luck. Move me along and I think we'll get nicely through our agenda.

Just to give you an idea of the topics that we'll be covering this morning, having been very efficient and effective, and no small thanks to Denice's moving us along, there are a couple of points that are to be considered this morning from the reviews yesterday.

One in particular has to do with kind of putting out for discussion, maybe likely to be brief, around the two health literacy measure that were passed by group from CAHPS and from AMA. If there are any points of overlap, discussion, points of consideration about those two measures. And I
think you were going to put them up for us to take a look at.

So that will be early on in today's agenda. And then we'll discuss measurement gaps. Gaps basically that I think you would have like to have seen brought in context of our measure review but didn't make it, but are nonetheless worth considering for further discussions or internal considerations with NQF.

And Helen, who will be here around 10:00 or so, had asked us to also consider, and have a brief discussion around, risk adjustments. And especially this issue of how you account for community level factors in the context of the measure considerations. It's a point of, from what I gathered from Helen, it's a point of continual revisiting and consideration of how you fit that.

I mean I for one try to figure out how do you distinguish what is supposedly outside the norm, or the scope of,
responsibility for practitioners or for organizations versus those that are actually malleable or mutable in some way, shape or form and should be part and parcel of any kind of charge.

So that's kind of the general agenda. We'll close out with, I guess, some discussion of next steps. If we're efficient about it we might even get out a little bit early.

But I also, before we even get started on that agenda, I wanted to see if there are any residual thoughts, comments, questions, points of order, points of disorder from yesterday's discussion and voting and process that you might want to bring to our attention. Yes?

MEMBER EDWARDS: So if we had commentary for any of the measures we should just send it to you, Nicole? Okay.

MEMBER TING: So I really enjoyed the discussion yesterday. I would like to
recommend, though, as we implement the measures for inclusion that we look at the phasing a little bit. And I don't know how possible that is given that many of the questions are already validated.

But I find that the current language focused very much in the care setting and that I would like to see the language expanded to include more stakeholders, like health plans, because there are actually a lot of questions that are applicable in those organizations too.

And I don't want our executives to take the easy way out and say, oh, it's a care setting and then not address them.

CO-CHAIR ANDRULIS: I guess that's, in part, a question back to NQF, has this come up in previous discussions in other measures? About expanded use or variation thereof?

MS. MCELVEEN: Yes, I think most of our measures are really specified more for
a clinical/hospital setting. But with this being disparities, which obviously is much more cross-cutting, I can understand your concern, Grace, about making sure these measures in particular are applicable to other care settings outside of what is specified.

I think that measure developers try to develop measures in a way that are applicable to a broader audience. But I'll make note of that.

CO-CHAIR ANDRULIS: Jerry.

MEMBER JOHNSON: It seems that that question speaks to who's responding to the survey, or the instrument. In most of the ones we reviewed yesterday the respondent was the patient or the person. In some of them the questions did ask about leadership and the plan. But they weren't directed towards those persons and so I'm asking for clarification on that point.

That seemed to be the heart of the question. Who was responding to it, because
some of the surveys did try to speak to leadership and administration. But they were not the ones answering the questions, they were not the respondents.

CO-CHAIR ANDRULIS: So what ends up happening in a lot of those circumstances where you take an initial model and you think well that could be very relevant to other settings but it probably needs an adaptation to those other settings. It may not fit, it's not an immediate one-to-one fit.

MEMBER TING: Right. So for example right now, I know that CAHPS is field testing their health literacy study with one of our WellPoint plans, as one of the three test sites.

So I can see those measures in the future definitely having very direct, tested, validated questions that for patients, for example, questions such as are you asked about your race and ethnicity or language preference. That's definitely something that
we an translate to members.

Do you find that you get information that you need in the language that you need. That's definitely applicable. So I definitely see there are a lot of elements that we should be asking to assess whether our organization is culturally competent and providing the right services.

And I don't want the language to be the initial -- it doesn't apply to us.

CO-CHAIR ANDRULIS: Elizabeth.

MEMBER JACOBS: Everything was sort of a blur yesterday. But I'm pretty sure the community measure did not pass, right? That community engagement measure. So I think that's a missing link. Actually Ellen and I were talking about that. We don't really have any good measures of outreach to communities or somehow assessing value in the community of the healthcare organizations. So I think that's one of the things that's missing.

CO-CHAIR ANDRULIS: That might be
also part of the gap discussion. Other comments, thoughts?

MEMBER EDWARDS: And I guess, related to that, can someone just quickly review how people would become aware of the opportunity to submit? Because Ellen and I were talking about that earlier.

So I can see situations where there might be people who would want to come forth with measures who just wouldn't even know about NQF much less what they should be doing. And how do you broaden that communication?

MS. MCELVEEN: That's a great question. So recently what we've done is we have tried to stay more in contact and communication with the measure developers. And by measure developers I mean the folks who traditionally are aware of NQF and submit, typically, to our projects. But what we've done is we've held webinars with them to keep them abreast on our process and answered
questions regarding the submission process.

The other tool that we have online is we allow anyone to submit a measure at any time, regardless if we have a current project for it. So they have an opportunity to start a measure submission or to begin that process whether we have a project for it or not. That will allow us to, number one, to be aware that a measure is available and give them an opportunity to submit that information.

I think there's still an ongoing outreach on our part, certainly to reach other groups and entities that are not aware of NQF and that are striving to, or that have information that could be useful to us. Do you have any other comments on that, Robyn?

DR. NISHIMI: Yes, I mean we also depend on you, you know, and Dennis's listserve and others. I mean we do broadcast when there's obviously the specific call for measures. But generally speaking I think more and more people are frankly aware of NQF.
What they're probably not aware of is that they can submit measures at any time now. That's relatively new.

CO-CHAIR ANDRULIS: Did you get any feedback from folks who might have been more typical submitters of measures about not submitting measures? Did anybody say, well, you know, we don't want to submit this time because?

DR. NISHIMI: No. And we usually don't get that kind of --

CO-CHAIR ANDRULIS: You don't?

DR. NISHIMI: Yes.

CO-CHAIR ANDRULIS: Okay. Try and move on to looking at the Health Literacy.

MS. MCELVEEN: Well, I'm going to do a short recap of yesterday I want to go through.

CO-CHAIR ANDRULIS: Okay, good.

MS. MCELVEEN: So again, I just want to thank you guys again for yesterday, we got a lot accomplished. And I would like to -
- first, one housekeeping thing. The thumb drives, if you used one of our thumb drives for materials yesterday, we just want to make sure that we got them all back, because we are missing one or two. So not that we're intentionally calling anyone out. But if you have one just remember to return it.

And so I'm just going to do a short recap of what we went through yesterday and then we'll start on the discussion around the related Health Literacy measure. For this presentation we'll direct your attention to the two large TV screens to the right and left of the projector, so you know.

So we reviewed 15 measures yesterday. One measure we still need to consider is the Cultural Competency Implementation Measure. That was submitted by RAND late, it was submitted this Monday. So we're going to schedule a future conference call with the committee to review that.

Out of the 15 measures, the group
recommended 11 measures for endorsement. Measure 1821, the measure focused on patients receiving language services supported by a qualified language service provider, we have noted that that measure is recommended pending the inclusion of that footnote to cite the Joint Commission and Office of Civil Rights references.

There were four measures that were not recommended. Those were two of the CCAT measures, one was on data collection and the other on community engagement. And the other two measures not recommended were from the Speaking Together program at GW and that was addressing patient wait time for interpreter services and the percent of work time interpreters providing interpretation.

So our next discussion, as Dennis mentioned, that we wanted to bring to the group is around a related measure addressing health literacy. And these two measures the group did put forward for endorsement. And
that's the Health Literacy measure from the CCAT and the CAHPS item set for addressing health literacy.

And so typically in our process for addressing any related or competing measures a measure that has the same measure focus, or the same target population, would be considered a related measure.

And so because these two measures have that same focus of health literacy, what we did is we would like to bring these to the group and we highlighted what some of those similarities are between the two measures.

So as you all know, there's several questions that have been outlined in the two measures. And I tried to highlight what some of those questions are. And I'll read a few to you because the print is a little small.

So for example, in the AHRQ measure it asks, "In the last 12 months how often were the forms from the provider's
office easy to fill out?" In the AMA measure it asks, "Were the hospital forms easy for you to fill out?" That's one example.

Another is, "How often were you offered help to fill out forms at the provider's office?" And then the similar question is, "Did the hospital staff offer help to you to fill out the forms?"

And then the last question that we found that was very similar was, "How often were the instructions about how to take medications easy to understand?" And the other question is, "Do you understand your doctor's instructions? Did you know how to take your medicine?"

So those, again, are the similarities. While they're overlapping in certain areas they're not identical for the target population. So for example CAHPS is geared towards the patient population and the CCAT measure is focused for patients and staff.
So the questions for the group are, first, if you agree that the measures have the same measure focus. If you do agree then do you also agree that they both should remain endorsed? If that is the case then we do need a justification for that.

MEMBER JACOBS: My question is you're not proposing just choosing some items over another, right? Because then if you change the items you could un-validate the instrument, right. So it's like we'd either keep them both or choose CAHPS or AMA, right?

Okay, just wanted to make sure.

CO-CHAIR ANDRULIS: Jerry.

MEMBER JOHNSON: I guess I don't know why we need to choose one or the other or if we think they both work, not just endorse both. Their method of selecting the population in their surveys are very different. I mean one is, as I recall, requires that there had been a visit in the last 12 months. I mean it could be a hospital
visit, it could be a outpatient visit or whatever.

And then the other one is more directly focused within something that happened within the hospital over a period of the last four weeks or something. Is that not the case with the CCAT?

DR. NISHIMI: The selection of the population is less, to me, the determining factor. I think the thing to think about is obviously if they had both been purely on patients then they would be directly competing and the NQF rules, if you will, would have sort of forced you into making a decision about the two.

Because these are two different target populations, leaving aside the issue of how the patient pool is drawn, that, in my mind, would be a justification for not choosing one or the other. It's just that we are sort of required by the process to bring this to your attention and for you to make
MEMBER JOHNSON: I guess I'm just trying to make the case that we want systems and providers to assess the quality of their health literacy work. And I don't now why we'd want to just say the only way to do this is through one qualified NQF measure. It may be there could be two qualified NQF approaches to health literacy. I'm missing that.

DR. NISHIMI: Yes, well in other measurement projects we do require you to choose, because it's about standards.

MEMBER JOHNSON: I'm just questioning the wisdom of that, that's all.

DR. NISHIMI: Yes. That's a corporate position. To have a single standardized way, because these are different populations, I think, it's a different discussion.

CO-CHAIR ANDRULIS: Mara, then Donna.

MEMBER YOUDELMAN: So I guess I
would make the proposal that we do keep both. Both for the reason that, Robyn, you were saying which is the populations are somewhat different. And also given that CAHPS versus CCAT, who uses them in general, is also different.

So I think there's a strong case to keep both with a valid justification and I'm not sure if there's any disagreement in this room and maybe we can just, you know, if anyone disagrees raise an issue now and then we can just vote and move.

CO-CHAIR ANDRULIS: Donna.

MEMBER WASHINGTON: The populations are different. It actually goes beyond just looking at the patient versus patient and provider. Since one is at the healthcare organization level, that would be the CCAT one, and CAHPS will also be incorporated into MEPS from what I understand from yesterday's discussion. So that further differentiates the two.
CO-CHAIR ANDRULIS: Anybody else, because I think we're getting kind of a clear --

DR. NISHIMI: Let's put it this way, does anyone object to advancing both measures? Okay. Does anyone feel that there were any other measures, amongst those that you reviewed yesterday, that were related and competing and that we need to consider?

These are the two that we identified, obviously, in the list. But we sort of are compelled to ask you if there's any others you want to revisit because of this issue.

MEMBER JACOBS: Can you put up a list of the approved measures? Like I said, I think only two of them were not endorsed. Is that right? Four, oh, four.

CO-CHAIR ANDRULIS: Grace.

MEMBER TING: Okay. So yesterday was really full and maybe I was just dreaming, but were there two somewhat related to
clinical cultural competencies of an organization? Or is it just one? Do people, could someone help refresh my --

CO-CHAIR ANDRULIS: Jerry.

MEMBER JOHNSON: Yes, actually I'm glad you brought that up. Actually the CCAT had a clinical competency one too as well as CAHPS has a cultural competency one.

MEMBER TING: Okay, I thought so. Right. So if we could take a quick look at that. I mean, obviously, I think if we endorse both of these we probably will look at endorsing but I just want to make sure how similar or dissimilar they are.

CO-CHAIR ANDRULIS: Ernie.

MEMBER MOY: Mine isn't specific to that, it's just a generic question. I think it makes sense to weigh a measure versus a measure to see if there's one that's better. But both of these are measure sets. And so if you take out a measure from a measure set all the validation is no longer valid. So I
don't see how you break a measure set.

   DR. NISHIMI: Well we broke the set up yesterday when we didn't endorse it, so --

   MEMBER MOY: The whole reason for doing them separately is because they had separate reliability and validity testing.

   MS. MCELVEEN: So from the AMA measure set there was a cross-cultural measure and then we had a cultural competency measure from CAHPS, are those the two that --

   MEMBER JOHNSON: What was the number on the first one?

   MS. MCELVEEN: The cross-cultural? It was 1894.

   MEMBER JOHNSON: Okay.

   MS. MCELVEEN: We're pulling up the table now to see if --

   MEMBER YOUDELMAN: I mean one other thing that I can just mention as you're looking for it, to the extent that we all sort of discussed the CAHPS cultural competency
subset wasn't really about cultural competency as much as we would have liked. I don't know if there's as much of a conflict if CCAT really is about cultural competency. Except for the interpreter questions. But that may also be sort of another differentiation, is that even thought the titles sound the same that the actual content and topics don't.

DR. NISHIMI: Okay, does everyone have --

CO-CHAIR ANDRULIS: Donna.

MEMBER WASHINGTON: So it's true, I just pulled up the items for 1894, the CCAT measure, and they are more directly related to cultural competency, but those also had the weaker psychometric properties among the CCAT measures that were approved.

They only had three items in the patient survey set whereas they had a very robust number in the staff survey set. So the trade off is content versus validity.

CO-CHAIR ANDRULIS: Jerry.
MEMBER JOHNSON: Might be on this is the same as the other one. I think it's exactly the same question. I would recommend approving both of them. For the same reasons that we just went through before.

DR. NISHIMI: Is there any objections to that? Okay so we'll make note that the Committee considered these but affirmatively decided to push forward with both.

CO-CHAIR ANDRULIS: Yes, Donna.

MEMBER WASHINGTON: I think part of the rationale you can include the difference in populations.

CO-CHAIR ANDRULIS: Okay. We're going to move on to a discussion that was started by Liz about measure gaps. Gaps that were not closed or considered directly, or for that matter rejected in the context of our review. And I guess, Grace, you want to start us off?

MEMBER TING: Sure. And this
question is really maybe more for Robyn and Nicole. In the past has NCQA submitted measures or do they tend to keep their own measures because, you know, they sell it for accreditation?

MS. MCELVEEN: They submit.

MEMBER TING: Okay. So I'm actually kind of surprised that they did not submit this time around. Or maybe they did and it didn't make it in, because NCQA actually has a whole class multi-cultural distinction certification.

And I actually, when I attended training, there are some elements that I think is a missing gap measure. It's a process measure so I don't think it's really quantitative. But I actually think that it is a really important step in measuring an organization's cultural competency.

So specifically it's the element on programs. And so they set the measure, or the evaluation element, is having a written
program description for improving culturally
and linguistically appropriate services.

And under that there is a
community engagement component and what that
might mean or how that might be fulfilled. So
they say that an organization should have a
program description that includes written
objective, a process to improve, which has the
community element, measurable goals and an
annual work plan. A plan for monitoring
against those goals and annual approval.

So this is not something that just
because you have it your company is culturally
competent, but I think that it does speak to a
company's leadership commitment to a cultural
competency in a class. So specifically, under
the process to improve, they want to see a
process to involve members of the culturally
diverse community in the process.

And they said that this could be
met through elements for advisory panels,
community forums to review and solicit
feedback. And/or focus groups. So I think that that's something very concrete and people can say, okay, that's how you might engage a community for feedback rather than just say we looked at the census data and this is what we think.

So I don't know whether we want to reach out NCQA and encourage them to maybe look at this element. But they have many other elements that I think are a little but more cross-cutting than just the care setting too. And I just wanted to propose that.

CO-CHAIR ANDRULIS: Go to Mara and Dawn and then Ellen and Lourdes.

MEMBER YOUDELMAN: I agree with Grace. I wonder does NCQA generally submit its things here?

DR. NISHIMI: Yes, and they were part of the previous cultural competency project so they presumably chose not to for whatever reason.

MEMBER YOUDELMAN: That wasn't my
question though. And I may have just missed this. But for example GW's whatever, sorry. Wow. Thank you, measures, that -- Wow. GW's measures were tested in hospitals. When we endorse them does that mean the endorsement is limited to hospitals because that's where they've been validated? Or anyone can now pick them up and use it?

DR. NISHIMI: Anyone can use it but it's endorsed by us as --

MEMBER YOUDELMAN: A hospital measure.

DR. NISHIMI: -- the applicable care setting will say hospital. But it's not like others can't use it.

MEMBER YOUDELMAN: And we can't say, because of the validation that we received, we can't say this hospital standard is also applicable to clinics or provider's offices or health plans. We can't say that? Or we can?

DR. NISHIMI: That's --
MEMBER YOUDELMAN: That's beyond our scope?

DR. NISHIMI: Yes.

MEMBER YOUDELMAN: Okay, I just wanted to double check that.

DR. NISHIMI: I mean we can indicate that, in narrative, that the Committee also felt that it could be a appropriate, blah, blah, blah. But when you see its endorsement status you will that it's -- you couldn't just change it in that respect. We could try and craft narrative that indicates you thought it could apply broader. More broadly.

MEMBER YOUDELMAN: Then I guess I would suggest, and I don't if we have to go measure-by-measure, but I would suggest that we seriously look at that and determine that, because I think some of these will probably lend better to other settings than others might. So I don't know how to do that if you guys schedule it --
DR. NISHIMI: So that could be part of the follow-up that we send you, you know, something by email and then you would respond by followup.

MEMBER YOUDELMAN: Right, because I think would sort of help with some of the measure gaps. Like, some of this, if it's really only applicable to hospitals then there's just a gap just on practice setting. Not just a gap of we're missing a measure. But we're missing a measure in a setting.

So I think there's that duality. So I think a good chunk of measure gaps are just, most of these were developed in hospital settings. So we're looking at having them brought into other settings is a big gap.

DR. NISHIMI: You'll have the opportunity to review the report. But also in followup emails staff will query you as to what you think about the existing.

MEMBER TING: Right. And that really is a key concern of mine. I really
want us to look at healthcare equity from a much broader sense rather than just care setting, because I think that this area is so multifaceted you really need to engage all stakeholders. And I agree with a gap of settings, very much.

CO-CHAIR ANDRULIS: It may be worthwhile sending out for discussion, or for feedback, to you some kind of chart that kind of lays out the measures and then issues, broader applicability, issues related to broader applicability, other points that I think we could all comment on. Because I think many of us would be interested in that breadth of consideration. Dawn.

MEMBER FITZGERALD: I'm going to get out of the weeds and back into the big statement world here. So earlier in the comments when you were saying, and I understand being the pragmatist here about, you know, if there's two measure that are equally valid that there's sort of this give
and take and one has to pick a superior measure.

But you know it seems as though NQF is sort of treading into different waters now where we're no longer talking about clinical measures of quality, but perceptions of quality.

And I think it's going to be more challenging to kind of have that model. You know I'm sitting here thinking it's sort of like forcing someone in the industry to say is Lean Six Sigma better than ISO.

They both have the same measure domains but yet we select, because they're appropriate to our industry or our setting or for whatever purposes.

And I know you all have much smarter people than me involved in this process but it seems like there needs to kind of some conversation around how one evaluates culture things like that from a different perspective, because I think it's true that
they're more embraced within domains inclusively and not sort of being able to tease out and say well this question in this one matches this question in this one, now all of a sudden we think there's overlap.

And I don't know how to resolve that, but it's just a comment because it is kind of a new world in terms of the kinds of measures that are coming up these days for NQF endorsement.

DR. NISHIMI: And I don't disagree. I don't think it's really an issue for this project, at this time. But, you know, we have a maintenance process and when these measures come up next time there may well be seriously head-to-head competing measures and that's the kind of conversation absolutely --

MEMBER FITZGERALD: Well and I know there's a lot of hospital culture surveys that will touch upon every single one of these issues we move forward. There's a Culture of Patient Safety. There's Hospital Leadership
Survey.

All of these things are going to have elements that touch on at least one or more of these other elements. So how do you, at that point, given the comprehensive nature of what the assessment is, sort of decide which of this one works better than which of that one?

DR. NISHIMI: No, I agree.

CO-CHAIR ANDRULIS: And I also wonder whether it would be worthwhile somewhere kind of cross-walking what we have reviewed here with the efforts that NQF had done around cultural competency standards to see where there had been a match. I think that may yield its own gaps. At least, not necessarily in terms of saying, oh you have to come up with a whole bunch of measures.

But get a sense of where there has been some fit and some progress of grounding in measurement some of those. And also it links to other work.
DR. NISHIMI: I think that will come up when you see the RAND measure.

CO-CHAIR ANDRULIS: Ellen.

MEMBER WU: Mara, I'm surprised you didn't say something about this. With the gap measures. So I'm really concerned that we don't have a measure around data collection. It seems a fairly easy thing to get at, of whether or not people are collecting race, ethnicity and language data. So I don't know how we address that.

And it would be nice to actually look to see if there's anybody who has some health related quality of life measures to talk about. To just get beyond the specific health conditions, but broader. Those are the two areas that I feel like there's a gap.

DR. NISHIMI: I think we do have some of those in the Patient Reported Outcomes project. The quality of life.

MEMBER YOUDELMAN: And data collection, though, I think that's Mara and I
are saying, there is an endorsed measure from
the last round that's specific on data
collection. It's not the evaluation of the
data collection, so that might be the
distinction we have to make, because I think
in the last panel, it was mostly preferred
practices, but I thought we adopted the HRET
Tool Kit as a measure.

So we have it, but I think what I
would say, from what you said, is then we have
to sort of get beyond the collection then to
the measurement of the collection, which the
CCAT was doing but maybe we need to figure out
ways to improve on the CCAT because that one
didn't get approved by us. Right?

MEMBER WU: I just can't remember
that about HRET has --

MEMBER YOUDELMAN: I'm sitting
next to her so it's --

CO-CHAIR ANDRULIS: Lourdes and
Grace, then Marshall.

MEMBER CUELLAR: Along the same
lines as what Grace and Mara were saying, we have a great opportunity now, there's 32 pioneers, ACOs that have been established. There's five in California, I know there's two in Texas. A totally integrated system to see, using the common medical record, does this continuum of information follow the patient all the way through.

And there would be more coming I think starting in July, I think it's the second phase. So with facilities I think that's a great opportunity.

I also didn't want to lose the point that Jerry made on leadership, because I think, just like CMS and Joint Commission, whole accountability at the highest level, being the Board of Trustees, with certain specific questions, I think, all the way to the Board and the C-suite, I think it's incumbent because they're the leadership.

I mean you can have all the training you want, get all the information you
want, but unless that leadership dictates it down it's not going to happen. So I think we've got to hold them accountable to a certain degree as well.

CO-CHAIR ANDRULIS: Thank you. Grace.

MEMBER TING: Thank you. So again, not to tout NCQA as a possible option again, but NCQA does have an evaluation on an organization's ability to collect and use race and ethnicity and language data. So that's actually their first element. And they do have scoring tiers that says, okay, if you do this then you get 25 percent. If you do this then you get 50 percent, 75 percent, 100 percent.

And it's a combination between patient directed -- collected, sorry. Race and ethnicity and language I think, based on OMB standards, as well as the use of indirect methodology, which is predictive algorithms using that as a quality improvement tool.
So I think that, again, since these are gaps maybe we're not asking NCQA, saying give us your whole tool. But just say that we know there are gaps and the CCAT one wasn't really appropriate could you consider submitting these, where we have gaps.

CO-CHAIR ANDRULIS: Marshall and then Romana.

MEMBER CHIN: So I just want to raise this issue under measure gaps. I think because of across, I guess, probably the two days and it has to do with, I guess, not so much our messaging and our language. So that the past day, and quarter, we've been talking about cultural competency and sort of like non-disease specific measures of communication and literacy. So fairly general. And this afternoon we may want to talk a little bit about these disparity sensitive measures.

But I think it's critical in the final product that we don't have as clean a distinction in terms of, you know, as I'm
hearing it I think someone could interpret as well disparities measures really are, it's cultural competency, it's literacy, it's communication. And then this more traditional sort of process in outcome measures that are stratified by REL or SES or something else.

So I think that we could misinterpret it. So I'll give you a specific example. I'm part of another NQF Committee called the Measures Application Partnership, which is devising which measures are used for public reporting and incentives and all.

And if you picture sort of a committee like this where you're the only disparities person, and so this is a general point that I've been raising at each of these different meetings, and I'm not sure how much it gets through, because actually they're looking for this group, probably say well we've come to disparity measures we'll look into this group, us here, in terms of what the answer is going to be.
But there's sort of this -- you can tell it's sort of like tinged in terms of what they're thinking about. Well, you know, it's say Marshall likes disparity measure. And so, again, I think it has to do with the messaging we're doing but also it dovetails with the discussion we're going to have this afternoon. But how is this going to be done in terms of disparities measures. Measure gaps, messaging, what goes forth in this committee?

That's the point I want to make, is then I don't want us to lose then, sort of, these two different components because I think they're complimentary but they are somewhat different in terms of what we've been talking about the past day. And then the usual sort of clinical measures stratified by some factor.

It seems sort of implicit I think, in the materials you've been sending, us that you guys, you know, at NQF may not be making
this just, maybe distinguishing so that I think people can potentially misinterpret if we aren't careful with how we package this.

DR. NISHIMI: What do you mean misinterpret what --

MEMBER CHIN: Well for example like this MAP Committee, they can say okay, disparities measures, we're going to focus upon the cultural competency, literacy, communications measures. These are the disparity specific things. And the RDF, like well looking at the other sort of usual measures that they are looking at but stratified by race/ethnicity, somehow different.

I think we get it. But I think others that aren't in this area, you know, they hear disparities, I think really it's a danger of us missing the boat.

CO-CHAIR ANDRULIS: Yes it's the difference between perceiving them as general measures that can be stratified versus ones
that are specific to issues of race, culture and language.

MEMBER CHIN: Right I mean philosophically it's this issue of like you think about your pillars of quality and how equity, I mean it's gone from the IOM 6, I think it was Number 6, to the newest IOM iteration, equity cuts across all of now seven pillars of quality.

And so it's a philosophical change that I still think that there are many in the outside world sort of view it as sort of something that could be marginalized as opposed to really being an integral part of everything we do.

So the idea is sort of like someone could misinterpret this and say, well okay, cultural competency, that's disparities. You know, we address cultural competence that's it in terms of our disparities efforts. So again, this is an issue I think we should try to avoid.
CO-CHAIR ANDRULIS: That's a point back to NQF about distinguishing and making sure that the issues are on race, culture and language that we're taking up around cultural competency, et cetera, are distinct from these other ways of stratifying data on race, ethnicity and language.

MEMBER CHIN: In other words I'm here on disparities measures but we've been talking about the past two days one critical set of components, but there are sort of critical components that were drawn on the same level that maybe we can't sort of put aside in terms of when others think about disparities.

CO-CHAIR ANDRULIS: Thank you. Romana, then Mara and then Donna and then Francis.

MEMBER HASNAIN-WYNIA: So one of the things that concerns me is, you know, I don't know how to put it without sounding negative, but this notion of kind of a message
glut and who actually adopts the measures that we endorse and that NQF puts out.

So I worry that there's this tension between kind of pushing the field and wanting the field to go down a specific path because we all believe in it and we all believe equity should be part of everything we do in health care.

But I also worry that if we have too many measures and the measures have a lot of questions still up in the air that we may just get this kind of, oh that's another NQF measure. So that's kind of one point that I want to make.

So I'm a little bit concerned about that and it was actually Allen's comment that made me think of this because, Allen, because of your comment about the gap in REL and not having a measure, at this point, that we've endorsed around data collection. And those of you who know me know that I've done a lot of work on data collection.
But I actually wonder is there a gap? And how are we defining the gap? Is the gap that NQF, that this committee, didn't endorse a measure on race/ethnicity data collection? Because from my point of view the gap would be in that NQF isn't aligned with, right now, kind of a little bit of a, you know, maybe a mini wave of endorsements around data collection.

So it's in the ACA, Meaningful Use, Joint Commission, NCQA, it's in the field. So yesterday when the AMA data collection measure wasn't endorsed I was thinking it's still going to happen. It's still going to happen whether NQF endorses it or not. That's something that's going to go forward.

So to me, the question is do we want to put out a set of measures where we are really pushing the field because we know that there's a gap and we know that the field isn't quite going there yet. And we want to put out
a set of measures that are strong and do get adopted versus saying, well there's a gap in what we've endorsed through this Committee.

But from my vantage point I don't see the gap. I only see it in, kind of in this room, that we didn't endorse it therefore NQF may come across as not being aligned with what's being supported by the general policy community and the accreditation world and so forth.

So I'm thinking out loud so if I'm kind of rambling, excuse me. But those are the thoughts I'm having in terms -- And I'm using data collection because I believe that we do need the data as kind of a foundation. So I'm a strong advocate of it, so I'm using that as the example. So what's the balance there? How are we defining a gap?

MEMBER JACOBS: Do you have a proposal?

MEMBER HASNAIN-WYNIA: Do I have a proposal to --
MEMBER JACOBS: Yes, how do define the gap?

MEMBER HASNAIN-WYNIA: Well I guess my question is are we defining the gap narrowly within the context of our discussion around a measure that we didn't endorse? Or is there a general gap in the field around activities not taking place that should be taking place? So that's why I used data collection. We didn't endorse a measure here around data collection. But I think data collection is going to move forward in the field.

But then that's a different, that's something that there may not be, you know, kind of another push outside of what NQF would endorse in the field. So maybe that's kind of loosely right now that's what I'm thinking.

MEMBER YOUDELMAN: It seems there can be different categories of gaps. And this wasn't going to be my comment but maybe it's
just fine. So one is stuff like data
collection or that NCQA has adopted, which
some of their stuff has actually made it into
their full accreditation for health plan. So
it exists, it's out there. It's just not
endorsed by NQF.

The second set then is sort of
what we talked about earlier, which is we now
endorse something for a hospital but there's a
gap because it's not for the broader provider
arena.

And then third, I think there's
categories of real gaps. So like I was
actually going to raise, we looked at cultural
competency in sort of a pretty narrow frame of
pretty much race/ethnicity language with a
little bit of literacy.

But disability, LGBT, is not
something here. And so that, to me, is
actually a pretty significant gap moving
forward. So I think there's different aspects
of it. To me I think I agree with you that if
it's already out then it's going to be done, that's not a gap where I care as much about focusing.

I'm more concerned about the ones where it's either not hitting certain providers that we need to hit, would tie to the standards that we've developed. Or going beyond that to other populations that we haven't covered. But that's just my opinion.

MEMBER TING: Yes, let me add gender to the mix.

CO-CHAIR ANDRULIS: Donna. And then we'll go over to this side.

MEMBER WASHINGTON: Yes, I just want to pick up on comments that a lot of people made. So sort of tying it together when Marsha was speaking, and now when Romana's speaking, I'm just thinking about sort of a framework for thinking about disparities and maybe that's one way to sort of couch both the NQF endorsed disparities measures as well as our definition of gaps.
So first thinking about, sort of commenting on what Romana just said, there's other standards, which are sort of ideals, and there are the measures. And so we all agree data collection is essential.

And there are multiple standards that support that. But it would be, if NQF doesn't endorse, you know, we didn't endorse the measure yesterday and who knows if there are other measures out there.

I mean I think that that's one way to sort of hone in on what at least the gap in the measures are. So it's not saying that it's not important, it's actually sort of supporting one of your earlier comments about not flooding the field with NQF endorsed measures just because there is a gap or there is a need to measure something to assure that that standard is met.

So I think maybe in presenting sort of the disparities measures or the concept of measuring disparities as a whole it
might be useful to present this framework.

And then also, following up on what Marshall said about the disparity, I wonder is there a ranking of NQF endorsed measures? Like some are considered more important than others?

It just seems like the ones that are measures that we'll be discussing this afternoon that were endorsed for other purposes, but that would be the ones that would be stratified by race and ethnicity or LEP or other indicators, because those are high disparity measures, might be ones that would be even more important to help plans and others to adopt.

And so thinking about how to present NQF disparities measures, I would think about those things as sort of outcomes. This is what we're trying to achieve. These are sort of more proximate measures of health.

And then the cross-cutting things that we discussed yesterday as like here are some of
the explanatory steps that lead to those differences in outcomes so that people understand you really need to look at both.

CO-CHAIR ANDRULIS: Do you want to respond question about priorities?

DR. NISHIMI: Just in terms of priorities, no. We don't do that ranking. That's based on those who implement. So for instance the group that Marshall mentioned, the Measures Application Partnership, I don't know if you're going to rank in the future, but they make recommendations to CMS on what measures to use in certain programs.

But within the performance measures, the sheer endorsement process, there is not a weighting or ranking of better or worse. It's endorsed or not endorsed.

CO-CHAIR ANDRULIS: Now when I hear the four comments that we just went through I think it may come back again, Marshall, to your idea of how we message what we're actually putting out in the context of
saying, okay, there are gap, we recognize, these are the measures that were submitted. However, we also acknowledge that there are these other very high priority areas that are moving ahead.

Nonetheless, we are moving also through these measures to advance the field, yet again, that we acknowledge gaps in terms of these measures that were submitted, but not in terms of where the field is going.

There's that really important balance to make sure that what we do here is credible and relevant. Otherwise it may be seen as kind of tangential to some of the direction that the field's moving in. Francis.

MEMBER LU: Yes, I put my flag up several minutes ago, but I think many of the comments have come up that I wanted to say also. First of all the race/ethnicity data I think is a bit of a gap in that people might ask well why wasn't that part of the NQF
package of measures since this has been put forward by a number of other groups mentioned here, NCQA and the Class Standards and JCAHO probably and other places.

So that gap I think will be fairly glaring. People will ask why. And then the issue about the missing groups, like the LGBT and women and et cetera, disabilities. I think that I understand the nature of our focus here was focused on racial/ethnic minorities but, as we all know, disparities go beyond that particular lens, even though that's perhaps the major focus of the federal government, et cetera.

But again, as we know in the disparity reports AHRQ they have been reporting about disparities related to other categories of cultural identities, such as the ones we've just talked about.

So I don't know whether this is an additional project, maybe, that we might to recommend to NQF to focus on in the future or
maybe as part of this package of measures. I think we need to explicitly say that this has been the focus of this project and there are disparities related to other cultural variables that we would like the field to think about or to put forward or something along those lines.

So it's not inadvertently kind of stated or understood that this is the be all and end all. Do you see what I'm saying? I think that be very important to make that very, very clear.

And then the final thing was the issue of applicability of our measures. You know these were tested in certain systems but to what extent, is this kind of generalizable to all healthcare systems, you know, is that really clear? And should that be more explicitly stated? I don't know, I just put that out there.

CO-CHAIR ANDRULIS: Thank you, Francis. Ernie.
MEMBER MOY: I had some reflections on the comments that Marshall made. And I do think it's important to have some kind of topology for disparities as it were, so that we don't get mixed up in terms of our measures. And I think what I see is something that analogous to like the Andersen-Aday model, because you kind of have potential.

Basically you have potential access and realized access. Potential access like insurance and stuff like that. And then realized access is actually getting the care that you need. And I think disparities is something along that line as well.

So we have potential risk factors for disparities, which is what the focus of this conversation has been, like literacy and not having culturally competent providers. But then there's a whole realm of realized disparities which is actually looking at clinical measures and seeing that there
actually are differences across race and ethnicity.

And I am concerned that with those emphasis being placed on these disparities risk factors that someone can say, oh well we did the survey of cultural competency so we're not going to do the hard thing of taking our clinical measures and stratifying and looking at the realized disparities.

And so I think there's some value to creating some kind of topology and saying you should actually look at both. You should look at some risk factors. You should look at the actual realized disparities because that's important as well. So I put that out there.

On data collection I wish that we had a data collection and race/ethnicity measure yesterday, because I think it's really important. But then listening to the conversation today I think I appreciate that if we put out a recommendation just about collecting race/ethnicity and not about all
the other components of culture, then that might not be serving things well either, because there are a lot of other things that are important for cultural competency beyond race and ethnicity.

CO-CHAIR ANDRULIS: Colette, then Mary.

MEMBER EDWARDS: My comments kind of tie into, probably the most with what Marshall had said, and Romana, in talking about messaging and also the concern about having a lot of measures out there that are basically doing the same thing.

And I remember in the very first face-to-face meeting we talked about the issue of harmonization and having messaging to the people who are developing that they should be looking to see what other people were doing so we didn't get into a situation of them basically reinventing the wheel. And having a lot of measures out there as opposed to people focusing on new things or actually getting the
work done.

So I don't know how strongly that actually gets communicated. I just don't remember, to the people who are developing where that point is strongly made in terms of looking to see what people have already done that would basically address what they may be developing, because then there can be some statement in the messaging about harmonization. So that would kind of address what Francis was saying about it being a glaring gap.

It's not a glaring gap because we recognize that it's being covered someplace else and would also message to the people who are developing what we're really interested in is something new, different or focusing on getting things done as opposed to a replay, with a tweak, of something that's already out there.

MS. MCELVEEN: So I think we're starting to get closer to how do we do it is
the question. And is there a way to identify a document, some type of introductory something, that includes the breadth of this conversation. They're all important points. And we obviously can't include everyone's standards in an NQF document. But there seems to be a need to have some capsulization of the various touch points.

I mean, there's the Office of Minority Health stuff. There's the stuff from Joint Commission, there's the Class Standard. So there are a variety of players who have made contributions to look at cultural competence, cultural diversity and can we cite that richness of that body of work some kind of way, almost in its entirety. I don't know that we ever get to the entirety, so that indeed we have given voice to recognizing that there are lots of players at this table.

CO-CHAIR ANDRULIS: So yes, Francis and I sit on the Class Group as others have, and this issue has come up in Class as
well, that okay so you're doing all this work but how do you contextualize this in some way that it's relevant. That people get a sense of oh, I see how this fits. I see how.

So it's the roadmap or at least a sense of relevance, to me. And I think it may be, as it is that a lot of these reports kind of important and incumbent maybe for us to take a look at an outline of what would be put into such a document.

DR. NISHIMI: I just want to make clear that NQF has, and I think it's somewhat new and Elisa can confirm, a standardized report format, which is not to say that we couldn't reference these things and they're all in line now and link to them.

But if you're looking sort of for a treatise that reviews all this stuff and actually includes it in the body of the report, my sense is that the report formats don't permit that.

CO-CHAIR ANDRULIS: Do the format
reports allow input into, within the context of the different segments of that report format, so --

DR. NISHIMI: I mean you will review this, but what I heard from you all was sort of a call for a section describing this and perhaps listing them and cross-walking them, those kinds of things. That is not the kind of thing that fits within the formats. Obviously we could reference the work of others and provide links to that. But that's the way the reports get shake out here.

CO-CHAIR ANDRULIS: Mara.

MEMBER YOUDELMAN: I want to build on that with a question or a suggestion. And I think one way to sort of address some of this discussion is to make sure that the report brings back the previously endorsed measure on data collection. And to say it's already been endorsed. Here it is.

DR. NISHIMI: Right.

MEMBER YOUDELMAN: It's not being
up for approval now, but that it set the stage for this work, to some degree, and to really show that history and to make that connection.

DR. NISHIMI: There is a section on related NQF endorsed. And so that would have had a basis in that.

MEMBER YOUDELMAN: I mean I think that is the point is if therefore, and the main report references to it, then I think we can capture some of this. And then, of course, if there's an away to, either in the section talking about the previously endorsed measures, or an appendix to just get to some of these pieces, I think that would be helpful.

I guess my other suggestions are, is I don't know if it is worth a conversation with NCQA to ask them about submitting some or all of the multi cultural healthcare standards that they have.

Since we have to have a conversation about the RAND standard, which
was submitted late, it might be an opportunity
to have that conversation with them and say
are there any pieces of this puzzle that you
might want to submit and it becomes a
secondary piece, if that's allowable. Or at
least to just find out why they chose not to.
Was it affirmative for some reason, or
whatever.

And then the last piece I think
this is picking up on what Marshall and what
Ernie were saying is, is there a way going
forward that as new measures are proposed, or
then come up for renewal at the end of their
initial cycle, to put in part of that process
an evaluation of disparity sensitivity. And
to really build it in to the entire NQF
process as opposed to what we're doing now,
which is add-on measures.

And there's still a reason for
add-on measures. But I think what we're
really sort of saying is it should be based
in, it should be part of this process.
DR. NISHIMI: And it is now. It just wasn't before. For instance in the cardiac care project, several measures that were up for re-review came in and they didn't have the section on stratification by race and ethnicity and disparities in them and they just sent them back to the developer and said, yo', these have been endorsed for six years now. You either give us this data or we don't review it.

So it is baked in now. It's just that we had such, obviously, a huge part of the portfolio where it wasn't initially baked into the submission.

CO-CHAIR ANDRULIS: Okay, Liz.

MEMBER JACOBS: I just want to go back to something that Francis and I think Colette brought this up. My understanding of this process is that we weren't told that we just had to focus on racial/ethnic and linguistic disparities.

I mean if we think there's a gap
in issues around LGBT and other issues I think we should actually bring those up. I mean, we don't have good measures but that's still part of disparities and cultural competency. So I would like to see us not leave that off the table.

DR. NISHIMI: Anything else?

CO-CHAIR ANDRULIS: Other comments? Questions, thoughts?

DR. NISHIMI: Very excellent discussion, thank you. You want to go to Taroon?

MS. MCELVEEN: Thank you guys. So now we are going to have a discussion with the committee, as Dennis had mentioned, around community level factors for addressing risk adjustment. And my colleague, Taroon, is here to start that discussion with the group.

MR. AMIN: Great. I know that Helen wanted to be here, so she'll be here in probably ten minutes. She's just finishing up a board discussion. So my name Taroon Amin.
I am a senior director here at NQF. I am working with Alexis Forman, who's actually in the back here, looking at an expedited review of all cause hospital readmissions.

We are in the process of actually voting for two measures that were recommended for endorsement. One measure that was looking at a hospital level unit of analysis. And another at the health plan unit of analysis.

And while the specific elements of the measure are probably not as relevant for the discussion today I wanted to give you a little bit of the context of the nature of what I'd like your reflections on today.

So the measure that we were looking at is a hospital level risk standardized rate for unplanned, all cause, hospital readmission following any eligible admission within 30 days of hospital discharge. And it was tested in All Payer looking at ages 18 and older. So it includes Medicare and 18 and older.
During the evaluation we received a great deal of comments from the hospital, particularly from the hospital community, but from a broad stakeholder perspective, that the particular outcome of interest in this case, hospital all cause readmission, had very much to do with the socioeconomic status of the patients under evaluation.

So the socioeconomic status of the patients had a lot to do with the nature of the readmission and the rate of readmission. And the hospitals that disproportionately treated this population would be at risk of a lower performance based on what many considered to be community level factors rather than hospital level factors.

And since the hospital level was the unit of analysis this raised a considerable amount of concern and comments for the broader community.

The steering committee considered these comments but ultimately decided that
adjustments of SES and a risk adjustment model was inappropriate, mainly because of the guidance from this committee and the guidance from NQF, in particular, on this issue of race and SES that including SES variables and a risk adjustment model would inappropriately assume two different standards of care.

However, they struggled with guidance on what our potential guiding principles going forward for this type of concern, considering that emerging research and previous existing research actually demonstrates quite a bit of a relationship between race and SES and hospital readmissions, for measures that are currently endorsed.

This measure, looking at the previous measures that were endorsed were condition specific, this measure would be all cause hospital readmission. So presumably, from these commenters, the effect would be greater.
So the Committee ultimately decided to essentially look forward into the future to consider potential hospital level adjusters or potentially community level adjusters that could be tested and used for this type of application.

That was sort of a consideration moving forward. That would be one area that we would kind of look to this group for some area of reflection of what would be the effect of using hospital level or community level adjusters in looking at risk adjustment for this particular cause.

The other recommendation that came from the Steering Committee, which is more in the realm of reporting, but also begs a little bit of discussion, is requesting, in display of this data, that hospitals be reported against like comparison groups. And one particular example was using disproportionate tier hospitals and comparing them against each other for this particular application.
So those are two particular areas that we'd look for some reflection from this group, particularly because of the high stakes natures of this area of measurement for use in public accountability and, likely, payment programs in the future for hospitals. This was one area that we wanted to get some thoughts from this group.

And there may be more that Helen would like to add to this discussion but I'll leave it there and submit that to the group for discussion.

CO-CHAIR ANDRULIS: Yes, this is a long-standing issue. This has come in context of severity of illness discussions decades ago. I remember this in the 80s being an issue, that -- I've got to get out the Geritol.

(Laughter)

CO-CHAIR ANDRULIS: I know from a public hospital perspective that this was a very sore point. That there were issues
around referrals that they had responsibility of that. And sometimes they had no control over, in terms of, like, ties to long-term care facilities.

And there are all sorts of anecdotes about some of the long-term care facilities because of their contractual obligations, were actually sent back to die in the hospital after they had been discharged. They were readmitted in a terminal condition and then, because the nursing home didn't want responsibility for that person's passing.

So it's treading in very worn, but very sore, territory in a lot of ways. And stratification by hospital type might be one of the points really to consider carefully in moving ahead. Ernie then Donna and then Elizabeth.

MEMBER MOY: Yes, I know old conversation. You have to risk adjust because the facilities you can use are too different. You don't want to risk adjust in a way that
you lose the information, so I think the traditional approach is stratification. Your suggestion of stratifying by hospital characteristics is reasonable but you might also want to stratify by community characteristics so you're comparing like communities.

MEMBER WASHINGTON: Well I couldn't have said it better.

CO-CHAIR ANDRULIS: Liz, then Jerry, then Marshall.

MEMBER JACOBS: I actually say if you do this you let hospitals off the hook. Hospitals should figure out how to provide the same quality of care to those patients. And I would say a lot of that readmission does not have to do with the individual or the community, it has to do with the way -- Because I worked in a public hospital for 12 years.

And we don't have good services for serving these communities and that if we
risk adjust we're not forcing them to actually address the issues, which is that they're providing lower quality care to these patients. So I actually say it lets people off the hook.

And what we talked about in our first session, when Joe and who else came to present the paper? Thank you. We talked about doing it two ways, showing the unstratified and the stratified.

I mean how you would decide what to do on that, but honestly if we're going to stratify and we're going to risk adjust this stuff away no one's going to do anything about this problem, which is that if you're Black, if you're poor, if you don't speak English, you've got worse care.

And it is somewhat about community factors, but it also has to do with the hospital and the quality of care they receive. And I know because I practiced in that setting for 14 years.
CO-CHAIR ANDRULIS: I think Liz raises a very good point. I would say also though that because you risk adjust doesn't necessarily mean that you're avoiding the issue. You're just acknowledging the circumstances.

Acknowledging the circumstances to the extent that they need also to be addressed, not to say that they're risk adjusted away. You know, this is a long standing issue that is both infrastructurally and issue in some of the safety net institutions for example. But at the same time it's a broader fiscal -- Anyway. And organizationally.

MEMBER WASHINGTON: Just to clarify. I wasn't advocating for risk adjustment, I was advocating for stratifying. I think that it's important to present both the overall unadjusted as well as the stratified results.

The problem with just presenting
the overall results alone, without stratifying, I mean in essence sort of risk adjusting by stratification, is that hospitals that do not have a disproportionate share of either vulnerable patients, or other patients with characteristics associated with some poor quality outcomes, are essentially off the hook and rewarded.

So I would be concerned about sort of the reverse problem. Not so much that you're holding minority serving institutions accountable, but more so that others get inappropriately rewarded. Particularly if they're in settings where performance is tied to reimbursement. So I would definitely do both.

CO-CHAIR ANDRULIS: Yes, right. So you run into the same issue around paper performance. Kind of trending you want to move towards those who will make you look better.

MEMBER WASHINGTON: Right.
CO-CHAIR ANDRULIS: And so it's a complex issue.

MEMBER JACOBS: I'm sorry, you're saying that this would lead to like cherry picking or people, explain to me a little bit more about how you feel --

MEMBER WASHINGTON: So I'm thinking about healthcare systems, for example, sub-pay for performance. If you don't stratify, if you just look at overall results, without some sort of accounting for differences in patient populations, then you may inappropriately reward hospitals that are better performing because of their patients or community factors.

MEMBER JACOBS: I'd like to respond to that though. Because the issue of these community factors are totally confounded by who are the hospitals taking care of them too? What is it, something like 80 percent of African Americans are seen in 20 percent of the hospitals, something like that. And those
20 percent of hospitals are disproportionately low performing hospitals.

And, like I say, it's like I think that people are sort of blaming the patient. And I think we're the healthcare system. Our job is to actually do better for them. So I'm a little bit concerned about, yes, there are these community factors but the hospitals should, I see we don't want to penalize them, I see what you're saying. I think we had this discussion about actually paying people for more complex patients. Maybe we should reimburse these hospitals higher.

But I also think that I just don't want to recommend something that would promote the status quo, which is that these patients, who also suffer for these community factors also tend to go to these hospitals that are very low performing. And we're not holding them accountable for that low performance.

CO-CHAIR ANDRULIS: Okay, one more, I want to get other folks involved with
this.

MEMBER WASHINGTON: So in an ideal world then you would actually use that to target resources toward low performing hospitals?

MEMBER JACOBS: Yes.

MEMBER WASHINGTON: Okay, so to use the VA healthcare system as an example, then they have a very complicated process for categorizing patients into different risk categories. And then hospitals are reimbursed based on a combination of performance and patient mix.

And so, for example, the patient mix, low income patients, they get reimbursed at a higher rate, for example, than higher income patients. Or homeless patients, they get reimbursed at an even higher rate.

And so that sort of levels the playing field. And then you can look at performance without that risk adjustment. So maybe that's too complicated to advocate in
other healthcare settings, but that's an example of how you can sort of stratify in direct reimbursement but without risk adjusting the results.

CO-CHAIR ANDRULIS: Marshall, then Kevin and then, Jerry you're next. Sorry Jerry, Kevin, Marshall. I know I'll get to Mara and Grace.

MEMBER JOHNSON: It's hard to know where to begin here because I think what we're really talking about is fixing the entire healthcare system of the United States, which is a little bit complex.

(Laughter)

MEMBER JOHNSON: I really do think that's what we're talking about. But I've participated in a number of discussions about this all cause readmission and I'd start by saying that, I mean, I do endorse all cause versus kind of disease matched readmissions, particularly with older adults where most of
these admissions and discharges occur.

So somebody comes in with heart failure, they get readmitted for some other reason, for a fall, but actually they went out of the hospital unstable. And hospitals and systems need to address the whole person, not just part. So that's where the all cause comes in. I do endorse that.

And I do think that health systems and health plans and hospitals have a responsibility for a continuum of care, including outpatient care and the whole transitions piece. And if some persons are in communities that are low resource, compared to others, then health systems have a responsibility to those persons too.

I mean certainly the public, and we of the nation, can't ignore those persons. And they need care. They just are going to require different kinds of resources. To a large extent I think we're talking about a
resource issue and how do we fund persons who require more resources for care than do others.

I'm in favor of not risk adjusting but I mean stratify and take a look at the persons that systems and hospitals care for. The community level versus a patient-centered approach, saying what kind of resources do I need or someone else needs, versus the community that I come from, I don't know.

I'm grappling with the best way to define community level in 2012 in relation to any particular hospital as a geographical proximity to where somebody lives.

And I find that extremely complex when I think about, at least, the city where I come from, from Philadelphia. And the neighborhoods that are close to the hospital versus a little bit further away versus farther away and what's the community and what's not.

So let's stratify, we have to do
it. I think to a large extent it's really a funding issue and how do we pay for persons and to provide the resources that are needed?

CO-CHAIR ANDRULIS: I think, Kevin, you've had your tent up. And then Mara and Chris.

MEMBER FISCELLA: One way of looking at disparities is to think about a mismatch between the needs of the individual and the resources of the system to respond to those needs and to look at that in a variety of ways, what are the financial resources, but obviously it's culture and linguistic and so on.

And the better that match potentially the smaller those disparities. The greater that mismatch the bigger the disparities. And that's the problem that's been alluded to is that people who are more disadvantaged tend to utilize providers, whether it's physicians or hospitals, that have fewer resources.
And so this problem is real and Medicare is, I guess, getting ready to deny payment for people who are readmitted for certain conditions within 30 days and there's data out there to suggest that hospitals that disproportionately serve African Americans and low income patients will be disproportionately penalized which will worse that mismatch between resources and providers.

And I think stratification offers a reasonable compromise in terms of still holding groups accountable, but accountable with groups that have comparable resources. It's not fair to compare one hospital who really doesn't have the resources for a highly developed quality improvement program with those that do.

And I think Rachel Werner has shown that using national data, so we have good empirical data on that.

The other factor that we need to keep in mind is that factors such as race are
proxies for worse worth health through weathering or accumulative disadvantage over the entire life span, which means, in many cases, that you're going to see racial differences in health care. In part as a result of those factors that are not fully accounted for by ICD-9 diagnosis adjustment, including greater likelihood of being readmitted for heart failure at a younger age or so on and so forth.

So I think there does have to be a balance, otherwise we're just going to be tipping that mismatch between needs and resources in the wrong direction.

CO-CHAIR ANDRULIS: I'm going to suggest that we kind of focus more on these, what you started down the path more directly on, it's community level factors that might play. It actually starts the play in the world of social determinates as well. So, Marshall.

MEMBER CHIN: So some ways we're
recapping sort of the diversity discussion we 
had, I think it was at our first meeting, when 
Joe and Joel Weissman came. And in some ways 
it gets to an issue of a change in perhaps 
what NQF should do. You know, if you take one 
extreme of well we're just going to sort of 
put the stamp of approval on different 
performance measures as NQF approves them as 
one, you know, far end.

Something in the middle in terms 
of like stratification where it's, well, a 
mini step in terms of how you use the data. 
Well use the data but then stratify.

I think a third way, which I 
think, again, Joe and Joel nicely did in their 
article, and I think probably I would 
recommend we think strongly about doing here 
also is, basically the next step in terms of 
recommendation that accounts for the 
complexity that Donna and Jerry and Liz were 
talking about.

So if this is going to be used for
accountability purposes, these measures, besides stratification, I mean there are other things to take into account. So for example do you reward based upon absolute attainment versus relative improvement. The issue of if you have under resource setting is there some system that Donna mentioned there are others where additional resources go to the under-resourced setting.

You know it starts getting into a little bit of policy but I think if you sort of say the issue of well, here's what we want to avoid, you know, the cherry picking or the making things worse in a situation, we're not advocating a specific answer but here are examples and I think they're papered to this well. You know, here are examples of ways that people have built into the system ways to safeguard against that.

To me it's more honest in terms of addressing the complexity. And stratification is a good first step but this goes beyond it.
And in some ways this may not be different than other issues where people are saying, well you have an issue of like under-resourced settings, this is not a new issue.

But especially if it's the means for accountability purposes, and you mentioned that once the measure is approved it could be used for any purpose. In some ways if that's not addressed that's probably equally bad as a risk you may feel in terms of putting your neck out in terms of starting a little more policy oriented.

DR. BURSTIN: I apologize for being late, got stuck on a board call. I think what we're really saying is we completely concur with what the paper said. The paper said don't risk adjust based on race and ethnicity. And we've stood our ground on that and concur with that.

I think what came up recently as part of this discussion, and it's really a question if you're not so much about the
readmission measure before them, and Taroon told you the recommendation was like hospitals should be reported with like hospitals, very much along the lines of what you’re saying, is really is there more work to be done here to understand if there is an opportunity to think about what Joel and Joe actually put in the paper, is when there are indications of when there is a community level effect here is there consideration of what those community level factors could actually be put into a model. Because what we’re really talking about is the measure itself.

A lot of those other things you just mentioned, Marshall, and others as well, are kind of outside the purview of the endorsement process. You know, CMS can make a lot of decisions about payment, others can make recommendations. The question is on the measure itself. Would there be, we continue to believe you shouldn’t adjust for individual patient level factors on race, ethnicity, SES,
et cetera. The question is is it worth thinking about what Joel and Joe raised about would you potentially adjust for the community level capacity to take those patients and really be able to help them?

In the case of readmissions it's especially important, just that if they are at community capacity for followup it may, in the measure at least being currently being used for accountability at the hospital level, how do you sort of factor that in?

I really just asking because we're trying to think about should we do more work here to really help a group like you thinking through what are those community factors. Are they things you would stratify on? Are they things you would adjust for? What's the science of even knowing yet of what those things would be?

MEMBER CHIN: But how is it really different in some ways? I mean we talk about individual factors but the community level
factors to me it seems like it'd be the same issues that Liz and Jerry and Donna have raised of, I guess, the issues of resource or as history, as Kim was saying, so that whatever else you call it, an individual race variable or an individual measure of community deprivation, say.

It's the same issue right? So you could come up with a better methodology for measuring those, but I think in some ways it's just skirting the issue.

CO-CHAIR ANDRULIS: You opened up a can of beans here. So let's see if I can get this right. Mara and then Colette and Mary. Now let's just go that far and then we'll continue on.

MEMBER YOUDELMAN: And I'm still trying to sort that out. But I guess part of what my question is, and so this may point to the need for more work, is how much is the hospital responsible for helping improve the community options versus just taking it as it
comes and therefore you absolve the hospital from some responsibility.

And I think that's a question that I'd like to see more delving into because at least with the ACA requirement on the hospital required conditions they do recognize that language might be a factor. And so they are willing to give some money to help with interpreting and translation at discharge if that's going to help prevent hospital readmissions just because of language barriers.

But if you don't have a rehab center or a nursing home where language services are in place then discharging that person with language services isn't going to help if the person then needs the community supports.

But that sort of goes into the how much of the responsibility is on hospital to help identify and develop the community supports that will improve its readmission
rates versus just you sort of say okay, there's nothing in this community for whatever reason.

So that's a piece I guess I'd like to see more focus on in figuring out how you develop it and is it staff at the hospital, resources in the community, partnerships, et cetera, et cetera.

CO-CHAIR ANDRULIS: Colette.

MEMBER EDWARDS: So I guess I would say, kind of going back to what Donna had laid out, which I think is also in essence, I don't know if I would go as strongly to phrase it as a recommendation, but really was a recommendation from Joe and Joel in terms of the differential reimbursement that I think that it may be outside our purview but if it's not I feel very strongly that NQF, with its heft and reputation, needs to make a statement in that direction because this is at a critical point where lots of people are making lots of decisions about
reimbursement schema.

And if we don't a stand now the people who are in a poor position, it's only going to accelerate very, very quickly and get out of control. And I think the opportunity is, right now, to put that out there for consideration by the people who are making determinations.

And kind of to Mara's point, the issue of the Medicare not reimbursing for certain things has had an impact in terms of hospitals with resources doing something about what they're doing internally plus also what happens after the person is discharged and what's going on in the community, because they're the ones that then lose money.

So follow the money trail is what I'm saying. And we need to make a statement about it now.

CO-CHAIR ANDRULIS: Mary.

MS. MCELVEEN: So is it that we need measures that hold whatever institution,
no matter where they're located, responsible for getting maximal patient outcomes and providing maximal care? And if that's the standard how do we measure that, no matter who's paying for it?

It is almost like you need a navigator, every patient needs a navigator, to help them get the best care in whatever facility. Short of being able to do that are there ways to create measures that evaluate that so is it quality of life and quality and care?

And I don't know the answer but somehow if you can come up with a way to evaluate both of those things I think you change outcomes. And one of the ways to possibly consider is looking at what's currently in the literature in terms of looking at evidence based outcomes. So that's part of where the science is in terms of quality.

CO-CHAIR ANDRULIS: Dawn.
MEMBER FITZGERALD: My mind's actually a little spinning so it probably won't be very coherent, but this is interesting because we've been doing a lot of research in our own state in looking at readmission rates across Tennessee. We've done analyses for each of what we call our metropolitan areas.

And it's ironic because the assumptions that we went into a priori about what we would find in terms of high readmission rates was not proven true. In fact, the lowest rates of readmission in our state are in Memphis, which has the lowest SES, highest racial diversity in the state.

And the outcome has actually been that where the high readmission rates occur are in largely rural, small referral hospitals that generally have some sort of a relationship with a hub hospital like regional hospitals that connect to Vanderbilt for example. Or Upper East Tennessee hospitals
that have a connection down to the university hospitals.

So I guess my point is is that we're making assumptions about the kinds of community based adjustors or considerations that we need to take in place. And the fact is if you look, even in the state of Tennessee, the answer to what the community based issue is is different.

In Memphis we have a large pocket and volume of high repeat utilizers who are, you know, it's less than five percent of the population but over 25 percent of the costs. In East Tennessee it's a more broad based network, it's not largely affiliated with any particular zip code or geography, it's just a lot of people that have no other resources available to them but a hospital care setting.

So I don't know what to say other than when you start to drill down and think about things from a regional perspective the easy answers that we sit around and talk about
here aren't what happens when you go down to a regional level and start drilling into the data and hypothesizing what the true causes are.

And I mean that was sort of this lesson learned about sitting back and armchair quarterbooking what the important issues are without actually looking at it from a community's perspective.

CO-CHAIR ANDRULIS: Ernie, then Romana and Grace and Jerry.

MEMBER MOY: I agree that the core issues, I think what you're trying to actually measure with the readmission, and I think that what we're interested in measuring from a quality perspective is the stuff that the hospital did during the hospitalization, and after the hospitalization, and how that contributed to the readmission rate.

And we're not interested in the other major driver, which is the underlying community admission rates. And so I would
suggest, if you want to adjust out these community effects, maybe you can just adjust for the underlying community rate of admission. And then that would isolate this kind of quality contribution of the actual hospital.

MEMBER ANDRULIS: Romana, Grace.

MEMBER HASNAIN-WYNIA: Right, and I'll be the first to admit that this whole public hospital setting is very, very foreign to my world.

But I wonder whether there's going to be any value in looking at some of the best practices facilities or systems, like the New York City Health and Hospital Corporation or the Jackson County Health Systems, that have done really well in terms of quality and yet practice in a very diverse, urban, disadvantaged community.

And what are their quality drivers that lower some of these readmission rates, and see whether there are factors that can be
teased out in terms of thinking what to incentivize and adjust, just a thought.

CO-CHAIR ANDRULIS: Commonwealth Fund has had high performing health systems effort for a long time that's targeted safety net organizations.

And what ends up happening a lot of times, programs like Denver Health have come out as a leading safety net organization. What ends up happening a lot of times though with these kinds of promising or model programs is that when you go to replicate them there's Denver Health and then there's Denver Health.

And it's been hard to tease out those broadly applicable opportunities. And I think it may get back to what you talked, Dawn, you take it down to the individual level and the circumstances, the sociopolitical community circumstances just are hard to match. But nonetheless, there are efforts that continue to look at and see if you can
tease out these promising aspects.

MEMBER JOHNSON: I might be able to understand what the options are a little better if we talk more directly to what some of these community level factors may be that we're even considering.

For example, some things that come to mind are zip code or census track. Or it could be resources within a given geographical area, such as primary care doctors or other providers, that sort of thing, but just what these community level factors are that we're talking about.

Persons have tried to measure social cohesion. You know what, just which ones we're talking about. But what strikes me is that, in contrast to the community level factors, when we go beyond the individual there's the family.

What about family in community level factors? When I think about readmissions, just in a very practical sense,
the question is, is the person who leaves the
hospital living alone. Is there really a
social network around that person, like
someone in the house if it's a frail older
adult.

And then the word family is
culturally laden too. Families mean different
things. Depending upon the notion of a family
back in the 1960s, based upon television and
for those who remember Ozzie and Harriet, that
kind of notion of the family is probably a
myth compared to family that I know of and try
to work with in West and Southwest
Philadelphia.

But, nevertheless, I would think
that family factors probably have more of a
bearing than community level factors on
readmission.

CO-CHAIR ANDRULIS: Yes, we worked
on a project where we actually established
just a crude couple of indices, one a social
deravation index and another one a child
well-being index, where we just combined available data from census and other sources and looked at those in the context of the way cities and suburbs were changing over time. And there was a lot of face validity to what we'd come up with as you looked at the cohesive aspects of those elements within the index, and then using them as a way to get a sense of what are the support systems and the status of certain conditions within communities that would then have implications for health and well-being. Liz?

MEMBER JACOBS: I'm just going to express one concern about this adjustment for the community context. One, I think someone else brought this up earlier, I think it is really hard to actually define what community is.

And then the second thing that I would say, and maybe Ernest wants to speak to what he means by community, but I think the
second thing is, as we know from all the work done at Dartmouth that continues to be shown over and over again, there are regional differences in healthcare that are just geographic in nature.

And again, if we are going to somehow give people, make the level playing equal by saying, okay, you're in this community where healthcare expenditures are twice as much for the same healthcare costs, I know that's not what you're talking about here.

But I'm just saying using that as an example then, again, we're just adjusting to allow the status quo to keep going on. I think it's important to know what we're doing for each population but maybe not to give people a pass for, but I already said that.

But, Ernest, I'd be interested in what you mean by community and how you would adjust by community. Because I was just saying, I think it's really hard to actually
define community and what that means.

MEMBER MOY: Yes, I know. I was thinking more from Helen's question, is more work needed. And I think the answer is yes.

I thought it would be an interesting experiment because if we are, putting aside community resources, which is really important, but if you wanted to measure community resources you wouldn't look at readmissions. There's a lot of other things you'd look at instead.

So if you're looking at readmissions, I think you really want hospital quality of care. And I think that one of the big drivers you want to then take out of that are the community factors, so that you can focus in on the quality of care delivered by the hospital.

And I'm thinking that maybe a proxy for those community resources is simply the underlying rate of hospitalizations for community to take out the geographic factor,
to take out the fact that in rural communities you're more likely to hospitalize someone so that they don't have to drive 60 miles, things like that. This's more of a research, I think, suggestion.

CO-CHAIR ANDRULIS: Kevin and then Ellen.

MEMBER FISCELLA: A couple things, I think it's important to keep in mind what our overall goal is and that's to have an impact on healthcare disparities, particularly those that are going to improve and narrow those disparities in health.

I agree with the discussion around the difficulty in defining community factors. I do think you could adjust for the composition of who's in the hospital, which would be different than adjusting for the individual patient level factor.

So for example, you could adjust for the median income, zip code of the patient who was admitted, or the percent Medicaid, or
the percent uninsured. You could even do race and ethnicity.

Really all these are various proxies, potentially, for the resources that the hospital has to care for that group in order to level the playing field.

But I think that that would make more sense than to get into the whole quagmire of the community itself and how to do that in an equitable way.

CO-CHAIR ANDRULIS: Ellen?

MEMBER WU: I understand the desired need to focus and what we're looking at and measuring. But I feel like we're losing this larger picture that the hospital is part of a community.

And community clinics and all these facilities originally grew out of a need within a community. And they're a part of a community. And I think that part of their responsibility and charge is actually to manage the health, not just care, in
healthcare, but the health of their patient. So I don't know how to do it. But to adjust away all of the community factors, and just have the hospital focus within that hospital four walls, I think we're not going to get at the disparities. And we're not going to hold our healthcare system accountable for providing more health and wellness than just sick care.

And Jerry joked about we're trying to change the healthcare system, but in all of these little pieces that we do around quality and coverage, the exchange, there are opportunities to start adjusting it, to start transforming it a little bit.

There's a window open to really start shifting the way we do things, either through reimbursement or how we measure things, what we look for. So I just think that that's really important.

I know it's hard, I know it's complicated. But I'm really concerned that if
we keep doing the same old, same old, and
being very siloed about it, we're going to
lose this opportunity.

CO-CHAIR ANDRULIS: Liz.

MEMBER JACOBS: This one comment
keeps coming back to me. But going back to
what Kevin said, I think that, because maybe
these hospitals have to do more to take better
care of these patients, and so if we should
adjust for these things and then that way
it'll equalize the playing field.

Again, it's not highlighting and
addressing the problem, which is that these
hospitals need more resources. Is there
someway in which this could be used to
highlight how these hospitals need more
resources, instead of just giving them the
same amount of money for the reimbursment, or
doing something like Donna was saying, I think
is an issue.

And I think back to being at Cook
County and a colleague of mine said to me, who
now actually is in the leadership there, which drive me nuts, said to me not every patient with diabetes can get a retinal exam every two years.

We're a county, we can't do that. And I'm like, that is the standard of care, how can you say that. But that's what happens in these places.

And I'm afraid that if we say, okay, it's resource poor and so we should not hold them to the same standard because they need more resources to take care of these patients, again, it just promotes this kind of way in which we give second-class care to these patients.

DR. BURSTIN: Thank you, that was a great discussion. It's as complex as we thought it was, I think, when Taroon and I walked in.

(Laughter)

DR. BURSTIN: And I think probably where we've landed to date is probably
appropriate, which is comparing like hospitals to like hospitals as being a recommendation for reporting, and not adjusting at the individual patient level, which we agree with.

I will point out, interestingly enough, and this isn't really just about the readmission measure although it certainly brought it up for us in a big way recently, there is actually a significant pot of money available through ACA, the Affordable Care Act, for hospitals who perform poorly on the readmission rates.

So some of this is also, you don't want to adjust away those differences and have the hospitals that are actually the least resourced to do poorly and not get that pot of money. So these are really complex issues so, thank you.

CO-CHAIR ANDRULIS: Okay, Romana.

MEMBER HASNAIN-WYNIA: To Liz's point, I was just telling Marshall, there's this very interesting little article, if you
haven't seen it, by Jan Blustein, in June 2010 of the PLoS publication, the open access, entitled "Hospital Performance, the Local Economy, the Local Workforce, Findings from a U.S. National Longitudinal Study."

The only reason I point that out is because one of the things that Jan does in this analysis is she looks at improvement and attainment.

And the thing that's very interesting is that the hospitals in the very under-resourced communities, after a pay-per-performance implementation, all improved. They don't all close the gap but their absolute improvement is far greater than any movement that was made by the hospitals.

We all know this, those of us who look at pay-per-performance and improvements. And so it comes back to Marshall's point in terms of should we be paying for absolute or for improvement. And I would just encourage whatever NQF puts out to --
DR. BURSTIN: We'll definitely pull that. The Values Purchasing program does, in fact, do that. It pays for both the actual attainment of a goal versus the journey getting there so it is interesting. Great, thank you.

CO-CHAIR ANDRULIS: Last word, Kevin.

MEMBER FISCELLA: At the risk of introducing what may seem like an irrelevant topic, I will say that with No Child Left Behind there's a realization that people needed to move beyond absolute performance, that is every child would be at this adequate reading level. I think it was by 2014 or something like that, otherwise you would risk being closed down and all these punitive sanctions, to really progress towards a goal and in looking at how a cohort of kids are doing and improving and finding ways to incentivize realistically attainable goals.
CO-CHAIR ANDRULIS: I think we're on for a break for ten minutes. See you back here in ten minutes.

MS. MCELVEEN: Yes, so what we're going to do is we'll take a quick break. We realized we just had breakfast, lunch is out because we originally planned for 11:15 lunch. It's a little early.

What we think is the better thing to do is take a break, come back, we'll start on the next piece and then break for lunch in an hour. Is that okay with the group?

(Whereupon, the above-entitled matter went off the record at 10:59 a.m. and resumed at 11:14 a.m.)

CO-CHAIR ANDRULIS: Okay, we're going to move on to a discussion around the Disparity Sensitive Measures Assessment. And there are a few questions that will be put forth to us for consideration. And for that I hand you over to Nicole.

MS. MCELVEEN: Okay, so you all
may recall one of the other pieces we're involved in is identifying measures that are NQF endorsed as disparity sensitive. And we have proposed protocol to the group that you have provided a lot of feedback on.

And so the first thing I want to do is just quickly recap what that protocol was and then discuss with the group the process that we've continued on and the results of that process in this assessment.

So now, if I can direct your attention to the large screen in the center, we'll use this for the slides following. So there were several pieces to this protocol, again, proposed initially through the commission paper to the group.

And so what was decided is that it would be separated into two tiers. The first tier is looking at prevalence, quality gap and impact. So specifically within prevalence, we've directed our attention around measures that address the following healthcare
conditions that are listed there.

So focusing on any measures that address, cancer, diabetes, for example, tobacco use, oral care, substance abuse, as well as cross-cutting areas such as patient safety care coordination, our palliative care and any measures around child health or pediatrics.

Second component is around quality gap. And within our measure evaluation form there's a section that specifically asks for information around disparities as it relates to the quality gap.

And we're using that particular section to identify measures and to fill in that indicator. And I'll talk more about that shortly.

Third, on the first tier is impact. And we're assessing that by deciding whether a measure can be mapped to any of our national priorities partnership goals or measure concepts that are laid out through
that work.

The second tier of the protocol focuses on care with a high degree of discretion. And to assess that, we're reviewing the measure submission forms that have indicated or cited a guideline as part of the evidence for that measure.

Second is addressing community sensitive services. And we are assessing that indicator based on if a measure can be identified or matched to one of our cultural competency practices addressing communication, or any practice falling under the care coordination project that addresses communication.

The third component is social determinant dependant measures. This indicator, the committee had quite a bit of feedback on at our last call.

But we're assessing this based on whether the measure is primarily within the direct control sphere of the healthcare
delivery or public health system, or whether it addresses a behavioral aspect of healthcare or is primarily an environmental aspect of healthcare.

And then finally we're tagging all of the measures based on a specific category that's laid out, so whether the measure is more focused on practitioner performance, whether it's indicated to be hospital or ambulatory care, home health.

And then also if it's a system provider based measure, whether it's cross-cutting, whether it's a structure process or outcome, so those indicators are ones that we are identifying for all the measures that are included in the assessment.

So to date, let me kind of go through some of the results that we have completed. So to date we've reviewed about 250 measures. Out of those measures 114 were included in this assessment.

And so in the review process of
our portfolio there are many measures that were previously endorsed that are now under review again as part of our maintenance process, as well as annual updates. So if that is the case, we did not include that measure in the current assessment.

So looking at prevalence, prevalence was very high, of course because we specifically addressed measures against it within those conditions that I just read. So about 94% or 80% of the measures that were reviewed scored very high for prevalence.

Looking at the indicator for care with a high degree of discretion, again, does the measure form cite a clinical guideline? About 60% of the measures were linked to a specific clinical guideline and a citation was provided.

Communication sensitive services was a little lower and that wasn't really a surprise to us as we looked through the measures. But there were really only five
measures that could be mapped to the practices that we've laid out within cultural competency and care coordination. And those were measures from the Child Health and Palliative Care Project.

Social determinate measures, majority of them were identified to be within the direct control of the healthcare delivery of public health System. So that was well over 100 measures that we went through. And then the remaining measures shook out. For process measures there are about 64, outcome, about 50 measures.

And then about ten, or a really small percentage, around eight percent of the measures were scored high for all of the indicators of the protocol. So that includes measures that were also linked to the practices as well.

The other piece of this process involved, again, identifying the quality gap. And on our last call I communicated to the
group that many of the measure forms did not indicate information around disparities. And so the goal to fill in that information was for NQF staff to do a literature search and to do our best to fill that information in.

And so the quality gap, in the large Excel spreadsheet that you all received, it includes a numeric value based on specific information that was included in the measure form.

So approximately 60 measures, or 50 percent of those that we assessed, we were able to retrieve that information, either from the measure form itself or based through literature searches that we did on the staff side.

And the distribution was pretty wide. It varied, as you can see on the screen there, from 1.5 percent negative, 1.5 percent to 39 percent. And we'll go through more details of those numbers in a minute.

This slide shows that distribution
in a little more detail if you can see the numbers. So the range is listed on the top of the graph. And then the number of measures that fell within that particular range is listed on the bottom.

So again, most of the measures were less than one percent for a quality gap. And then several others falling between two and three percent and then another third or so fell a little higher between ten and 20 percent.

So within the large spreadsheet that we provided to you there were also, of course, outliers within that gap information. And we tried to highlight those cells on the spreadsheet. We did highlight them in blue if you're viewing that. And we'll project that in a minute.

So more than 70 percent of the measures that we identified had a quality gap of ten percent or less, as I just mentioned. But specifically speaking to the outliers,
there were measures that were specified more on a population level. So, of course, when you're thinking about a quality gap for a larger population it was just a larger number, naturally.

There were also measures that were not specified in a percentage or numeric value. It was more of a narrative given to address disparities.

We also included quality gap information around the general concept of the measure versus the specifics of what the measure was measuring. So those were the three outliers that we wanted to bring to the attention of the workgroup.

So finally, the distribution for the scoring, as you know, for each measure that we tagged there was a scoring at the end that was provided. So the scores that are listed do not include quality gap. Because we did not specifically score that indicator because we were still working on it.
But just to give you a distribution of the scoring, for the first tier the scoring ranged from about three to five. And then second tier was one to nine. Overall scores were distributed between five and 13. And when we pull up the spreadsheet we'll be able to talk through some of the specifics around that and what that really means.

So with this information there's a few things we wanted to bring to the committee's attention and to get your feedback on. The first, and probably most important, is how should that quality gap data be used? So do you first think that you want to really consider the quality gap as a high indicator? Because we're really struggling with identifying that information and filling that in in a complete way.

Should that be weighted high in terms of tagging measures as disparity sensitive? Or does the committee want to
consider classifying the quality gap as, let's say, 15 percent or higher to really count towards the scoring of each of the measures?

And then how should we really address the issue of measures that there just was no data identified? And how will that weigh into the decision around identifying what the quality gap is for that group?

And then we do also have some additional questions around how to address the outliers and the scoring for the measures that we've identified, but first, taking it in pieces. I think the first step would be to start to think about that indicator of quality gap.

MS. MCELVEEN: We're pulling the spreadsheet up on the screen but you do have a copy of this in the electronic material that was sent out. So if you can't see it you might want to pull that out.

MEMBER WASHINGTON: Are you asking for comments now or were you planning to go
through the, actually a question and a comment. The question is how did you arrive at the 15 percent. That seems a little bit arbitrary in the sense that smaller gaps in high impact areas might be quite relevant so I would just --

DR. BURSTIN: It was totally arbitrary to get you to start talking about it.

MEMBER WASHINGTON: Oh, okay, great. And then just to start the discussion on how the gap data should be used, looking at the two tiers it seems that areas where there are either high gaps or some combination of moderate gaps and high impact, you might think about that as inclusion for considering that, or labeling that as a disparity sensitive measure.

While others you would then move on to the second tier and look at other factors. So I would not use either absence of data or lack of documentation of a disparity
as an exclusion, but rather as a reason to move on and look at other factors.

Part of the danger of taking the reported data, having gone through the process yesterday and having a better understanding about where some of the data comes from that's reported in these measures, then unfortunately I wouldn't make too much about the presence or absence of a gap.

It could be a very biased sample, it could have been targeted for specific purposes. And so it may not necessarily reflect the broader literature, on the one hand.

On the other hand the broader literature, which might document known disparities in an area, would point to the need for a measure. And it just may be a lousy measure for assessing disparities. And so it's useful information but it's not everything.

CO-CHAIR ANDRULIS: Nicole, do you
want to take us through a --

MS. MCELVEEN: So for those of you who are able to pull up the spreadsheet --

Sorry, we're just trying to choose the best example. So to highlight one of the outliers for, under quality gap again, we're looking at Line 77 in the spreadsheet and we've also pulled it up on the screen there. This is a measure from our cardiovascular project looking at hospital all-cause, risk-standardized mortality rates.

And so the quality gap specifically for this was fairly high compared to many of the other measures within this project. And that was around 16 percent.

DR. NISHIMI: So let's walk all the way across on how it was scored. It got three, the highest number of points, under the prevalence. Documentation was provided and the highest disparity was 16.8 percent. So that's the value there.

You see a blank slot because that
was where the committee was perhaps going to make judgements on. Zero to five is a small gap, five to ten is a medium gap, it might get two points. And ten and above are our example we just threw up for your consideration. Fifteen and above gets three points, again, remembering that there's a "total score" that's going to be at the end.

And if you keep going across you see it was assigned impact, a one, it didn't cite a specific guideline so it got "zero" points. This was matching to the cultural competency practices, is that right? Care coordination practices, things got certain points, the social determinant issue, that was a staff level judgement, is that right? Staff had to make the assessment there.

The next few aren't point values, they're just descriptive. So that at the end when you looked at the entire "disparity sensitive" set you could say, oh my gosh, we
have absolutely no process measures, or something like that. So that's a descriptive field, whether it was a consumer survey, provider base, et cetera. So those are all descriptive type things.

And then what was done was, if you recall what you discussed, that there was going to be a "first tier score and a second tier score." And so based on that cells were added up. And then there are various comments that we had to make to keep track of what was going on.

So that's what was done for each of the measures. And really, the question for you right now that I think Nicole wants to focus on, is this notion that we'll phrase the fact that, do you want that quality gap score to still be a first tier issue.

Frankly, given that we don't have data for 50 percent of them, you all might be able to point to a few more articles, et cetera, where we might have gap information.
But the sheer fact of the matter is, there are going to be a lot of measures in the NQF portfolio that don't have gap information.

And so then you have to start making qualitative assessments, you meaning you, not meaning us as staff, about how to weight these things to put them in or out for them to have the set narrowed somewhat for you to make some final recommendations.

And these are only what, half the measures, third the measures, that have been added. So we're trying to winnow the list down so that you can make more informed decisions. But we need to do so in a logical protocol-specific way so that what comes out at the end doesn't look to the outside like it was just this ad hoc, gut level thing.

MS. MCELVEEN: Kevin?

CO-CHAIR ANDRULIS: Kevin, you wanted to talk?

MEMBER FISCELLA: Yes, I think I
would advocate for considering the size of the quality gap. I look at this as very much a work in process. And we're just getting started here. And I think it makes sense, for a number of reasons, to start where we know disparities are and where we at least know them to be the largest.

That doesn't mean that there certainly aren't unknown areas that are much larger. And those will be identified through future research and then can be targeted.

I think one rationale for starting with where they're largest is simply the population impact. You're going to have the biggest, assuming you're taking into account prevalence as well, but if there's a bigger quality gap and you close that gap you're going to have the bigger impact.

Secondly, there's all sorts of statistical issues that come with looking at very, very small gaps. And whether one is really making a difference there, that becomes
more challenging.

And I think it's important to have some early successes and I think if you have a bigger gap that you're going after there's a bigger opportunity to show improvement.

CO-CHAIR ANDRULIS: Colette?

MEMBER EDWARDS: I have two questions and one is just so I understand the scoring a little bit better. Where it has care coordination practice, and I can't remember exactly how that was defined as not applicable, if I'm reading correctly. How was it determined to not be applicable related to death within 30 days and the whole admission.

DR. NISHIMI: Because it actually doesn't map to a specific practice. You'd have to have the report in front of you, unfortunately. Obviously it's a care coordination issue --

MEMBER EDWARDS: Right, that's what's confusing me.

DR. NISHIMI: -- which is a
discussion that we could have. But the way
the field was initially described, that the
committee viewed, was does it really map to
one of the NQF endorsed practices, the actual
practice, not whether it's a care coordination
issue.

MEMBER EDWARDS: And then my next
question has to do with what we were being
asked to do in terms of where it says, no gap
for identified and make a consensus decision.
Are you saying that we would do it measure by
measure or we would say, for all measures that
don't have gap data we're going to say they're
disparity sensitive? Is that what our options
are?

DR. NISHIMI: I think the decision
is at some level you all have to decide what
we do about those, measure by measure --

MEMBER EDWARDS: So it's open
right now, is what you're saying?

DR. NISHIMI: Yes.

MEMBER EDWARDS: I didn't know if
there was something implied by the way the question was raised?

DR. NISHIMI: No.

CO-CHAIR ANDRULIS: Liz, then Ernie and Ellen.

MEMBER JACOBS: I think the idea of going where we know there's a gap is a great one, like Kevin was saying, to help narrow it down. But I'm wondering if we also want to see, if we do that, are there important areas in which we want to do measurement that are left out, like is it all in cardiovascular disease or all child's health, or something like that.

And so maybe we want to do that and then say, okay, are there important areas that we're missing doing the measurement, and then add a few in that we don't know there are gaps but are likely to be gaps, to cover the breadth of things we might want to measure disparities or cultural issues.

CO-CHAIR ANDRULIS: Ernie?
MEMBER MOY: I think basically the same comment, but specifically I'm thinking how people might use NQF lists. And I'm kind of thinking that they might be a hospital and they're looking at hospital measures, or a nursing home looking at nursing home measures. You might take the measure within a provider type that has the largest gap and say that's more disparity sensitive than the others.

CO-CHAIR ANDRULIS: Ellen?

MEMBER WU: I think I'm just looking at this a little bit more pragmatically. Do you have it listed by gap measure? Can we --

DR. NISHIMI: By the quality gap field?

MEMBER WU: Yes.

DR. NISHIMI: Yes, she can do that right now, largest to small.

MEMBER WU: Yes, so then do we see a natural cut-off?

DR. NISHIMI: The point was you
don't really have, that's why we gave you that table of distribution.

That's the distribution now.

CO-CHAIR ANDRULIS: Donna, then, Kevin, you still -

MS. MCELVEEN: We're going to try and make it a little larger.

MEMBER WASHINGTON: We have to be a little cautious about interpreting some of the numbers. I just arbitrarily pulled up one of the cells to look at the details behind the gap. I looked at Number 1454, the proportion of patients with hypercalcemia. And the gap listed --

MS. MCELVEEN: I'm sorry, what line on the Excel spreadsheet is that?

MEMBER WASHINGTON: You know what, I sorted mine so I'm not sure. Look under Column A, it's measure 1454. Oh, there it is. It's right there, the one that's right on top. It lists a quality gap of 39 percent.

But if you scroll over to the
column with the explanation then actually it reports out the percentages by race/ethnicity. So for whites it's 39 percent, African Americans it's 41 percent, Hispanics, nine percent, and Asians, two percent. So that gap of 39 percent is actually looking at African Americans minus Asians. And I think there was guidance about how to calculate the gap looking at the historically advantaged group as the reference point. So in essence this 39 percent gap is actually a two percent gap.

DR. NISHIMI: Well actually though, the agreement when we went through the protocol was to choose the largest gap between, when we first reviewed this, so not just between the historically disadvantaged and not.

It was of the race and ethnicity data that we found what was the largest gap between the populations that were reported. We could change our minds now but I just want
to point out that that was the decision that was made and sent.

MEMBER WASHINGTON: Okay, but it still just reflects that whatever arbitrary cut point we come up with we also need to look within the data to understand what that cut point reflects, or what the data reflects.

DR. NISHIMI: Yes, but to me two percent of the people of my ethnic background, racial background, however you want to characterize it, are getting damn good care and other people are not. So I do think you want to use the largest gap, that's just me personally.

CO-CHAIR ANDRULIS: Kevin?

MEMBER FISCELLA: This is just a clarification question. The gap, does it reflect the absolute difference in rates between the highest and lowest? And these are always true rates that we're looking at the difference between?

DR. NISHIMI: For almost in all
cases, yes, but that's what some of the outliers were that Nicole pointed out. There was one that was based on there's a gap of point five nanograms per deciliter between two different populations. We reported that but you can't translate that to a rate. But when you see these percentage, those are rates.

MEMBER MOY: Yes, just on that so you might want to have special consideration for things that aren't percentages, so differences in mortality rates, which would be really low, which might still be important.

CO-CHAIR ANDRULIS: Marshall?

MEMBER CHIN: I guess another factor that maybe some of the different columns partially get at, but still maybe not immediately through the most logical end results, really have to do with the distribution.

The end results, like Liz was saying, you measure a scenario where at the extreme we have 50 percent of the measures are
cardiovascular or adult measures. I know that AHRQ has grappled with this somewhat in terms the same issue, in terms of you have different grids in terms of different factors, whether it's child, adult or preventive care, acute care, surgical, medical, et cetera.

But at some point there probably should be a check in terms of does it pass the SNF test of balance, some degree of balance. For example, I can imagine, say for example that there aren't a lot of child measures. And so child measures may not score as highly on these different columns.

But wouldn't want to have emphasis which we have no child measures. So some how that probably needs to be built in the system, some type of look at, are we missing major areas where -

DR. NISHIMI: Can we have some discussion around, because what we'd like to do is continue the screening and then bring back, to the committee, some of these cross-
cuts, if you will. So we would bring back any measure that was 15 percent or higher, whatever you land on. And you could take a look at that.

We would then do a sort of measures by area so that you could see that. And this could be all of them or it could only be those that are, let's say, five percent above.

A broader swath but of those that are five percent and above you've got 40 of them are cardiac, one child, one pulmonary and one ERSD or something. So then you could look at it that way.

We could do some of the other cross-cuts that you've talked about, but absent that kind of guidance, it does devolve to you literally having to go through line by line, which is what is Colette asked. We're talking about line by line because 50 percent of the measures don't have any gap information.
So the staff needs some specific guidance on, Number 1, even for those that we have gap information, how do we sort those and bring that back to you. But also, what are we going to do about these measures for which there's just no gap information?

CO-CHAIR ANDRULIS: Yes, Mara and then Ernie.

MEMBER YOUDELMAN: I missed the December call and I am just drawing a blank on this. But can you just go back to when something is determined disparity sensitive, what happens? Because as you said earlier, that going forward any new measure or reviewed measure is going to have to give this data. This is the interim process until everything is newer reviewed? I just want to confirm that, correct?

DR. BURSTIN: At this point now every time a measure comes up for maintenance we request that they submit the data stratified to look for disparities. So we'll
gather actual gap data on the use of the measure in performance.

DR. NISHIMI: But that doesn't mean it's necessarily disparity sensitive. You all are here to kick things up to that level or not.

MEMBER YOUDELMAN: So there's two different tracks then. There will be disparity sensitive measures, which must have this data. And there'll be reviewed measures, which, okay, maybe I'm just confused.

So reviewed measures will have to give disparities data but may not be disparity sensitive, which means what, if they're not disparity sensitive? Maybe I'm just completely confused.

DR. NISHIMI: So if there's no difference in disparities when it comes in through maintenance, then let's say it's a hospital measure. Then a hospital may or may not choose to take a very close look at it, which is not to say that there isn't
disparities within each hospital's own populations. But it creates a set where there's clear indication that this is a disparity sensitive measure.

DR. BURSTIN: And therefore it should be stratified.

CO-CHAIR ANDRULIS: Yes, it sounds like there are almost three groups you're talking about here. There's the group where there's data to show disparity, some differentiation that we want to look at.

Then there's the group that you have data that show no disparities. But we want to review that too, to look at the quality and the nature of those measures. So it also gives us a context to see which ones are showing up with no disparities.

And then the third group is where there isn't any information on disparities. So there may be all sorts of ways in which we'll cluster this.

But I think the idea of having us
have an opportunity to look at the way the measures that are currently available, along those three areas, can give us a sense of what's been documented, what has been documented but we want to review to see about whether the data stands, and then about those that there are no data, at least that have been found to date. Ernie?

MEMBER MOY: Yes, I guess I just hesitate to flag something as disparity sensitive in the absence of data. Maybe you could just call this pending or no information. And that way that would have people focus on those things that we have seen a demonstrated difference.

DR. BURSTIN: And just to add to that, I think that's absolutely right. And I think I do see it as two complementary processes.

So the idea here is to say, of the measures we have already got that are in hand, which of those are in areas where we know
there are disparities. They fit these
criteria, they should be labeled as such. And
we encourage people to stratify.

Then there's going to be the set
of measures prospectively coming to NQF where
we're going to be asking them to be submitting
their data on disparities and adding to that
quality gap piece, which we oftentimes don't
have.

And we'd like to eventually, and
with your help, think through how we get all
steering committees to look at those data and
make that determination, prospectively, as the
measure comes in.

They say either this is
retrospectively been assigned as a disparity
sensitive measure, wow, look, here's
stratified data on performance in the last
three years.

So I think we need to think about
but the retrospective piece of this and the
prospective piece of this, which I hope will
be complementary.

DR. NISHIMI: Right, but what I heard, and would be useful to hear, is that when there's no disparities information, that I heard what Ernie just pointed out. He would hesitate making any judgement on it. I also heard comments that they'd like to look at the full range and maybe pull some in there, even if it had no data. So I do think those are two competing ideas.


MEMBER WASHINGTON: So I've a question and then a proposal for moving forward. The question is, among the measures with quality data then what percent, approximately, were included in the measure versus gathered from the literature search?

And then as you're looking that up, the importance of the distinction, picking up on what I said earlier, is that even if a disparity exists the measure may not
necessarily be sensitive in detecting it.

And so I would more heavily weigh the measures that actually included the stratified data, or data on disparities in some form, in their submission.

So if you were looking to willow the list why not choose measures that have already achieved the standard that we're then suggesting for new measures going forward?

DR. NISHIMI: I can't give you a number. I do know there was either one or two, since I did some of the literature ones where they had actually used the measure and, for whatever reason, it was an older measure. It wasn't in the form when we required it.

So we were able to actually to get specific disparities information from the literature and plug it in because it was the actual measure, in essence.

But in terms of what forms, if we just do it by forms, half of the measures right now have disparities information. And I
think ESRD and cardiac maintenance was where
we really started requiring it. So it's only
going to be those two projects right now.

    Some of the others filled it in, 
kind of namby-pamby, so catch-as-catch-can, 
they provided it because before it used to be 
a field but not an emphasized field. So there
might be a few more. But systematically it's 
those departments.

    DR. BURSTIN: It's also very
dependent on the developer as well. So if you
think about it CMS usually doesn't have
difficulty with submittals. It was
interesting, they discussed it first but
managed to find all the data. And in fact
submitted it all in cardiovascular, mainly
because our chair sent every form back until
they submitted it, which was great.

    But some of other developers may
not actually have the data in their hands.
They've developed the measure but they don't
actually have the data. And so those folks,
it will be harder, even prospectively, having
them give us that data back.

CO-CHAIR ANDRULIS: Kevin, Ernie, Dawn?

MEMBER FITZGERALD: Is there any
relationship between what's the current score
now and whether or not there's data associated
with the quality gap? I guess I'm curious. Is
everything that's scoring really high, are
they the ones also that we don't have data
for, which would have me concerned about
saying you can't toss it out or is there any
relationship at all?

DR. NISHIMI: No, she's wondering
if the null fields have any relationship?

MEMBER FITZGERALD: Yes, in other
words, I'm looking at the range of score, the
total score.

DR. NISHIMI: Right, I understand
what you're saying.

MEMBER FITZGERALD: If it's ten, 11, 12, which appears to be quite high, but
yet those are all the ones we don't have quality gap then we've got a dissonance between what we think is a valid quality measure for which we don't have any data.

CO-CHAIR ANDRULIS: Any other comments?

MEMBER FITZGERALD: And if that were the case then I might argue for a position that says something to the effect of if all other relevant factors of what we think are disparities, related or high, and yet there isn't a quality gap present, then that might be treated differently for all other factors being scored quite low and not having a quality gap.

DR. NISHIMI: It's a mixed bag. Right now, the highest total score was 15 -- 13 sorry, I need new glasses, clearly. And then if you scroll over to the left for that cell there was no data.

But in the next one, just keep in mind that that right hand side is the highest
going down. So now look at the quality gap data and begin to scroll down. That's fine, she's still in there. No, scroll down, line by line.

Like I say, it's a mixed bag. Sometimes there's an actual zero reported gap on the form. A lot of nulls and some that have values.

MEMBER FITZGERALD: But is there not the capacity maybe to, and again, I'm not hopefully quantifying how much love there is in the universe.

(Laughter)

MEMBER FITZGERALD: But you have to take the factor of how many people divided by -- but seriously is there some way to put those two together? And saying that if all other factors of what we conceptually think of as quality are in that concept of value of ten or higher, then the lack of a quality gap wouldn't necessarily toss it out.

But the recommendation would be
that you need to present that and this update versus those that we scored low on all of those other factors and no quality gap, we'd say, forget it.

DR. NISHIMI: That's an excellent cross-cut. Any other thoughts on this torturous spreadsheet? And hat's off, really, to Adella, I'm talking about this and Nicole has been -- Adella has been --

CO-CHAIR ANDRULIS: Sean?

MEMBER O'BRIEN: I don't have an answer to what to do with the measures evidence, except to say that when I glanced at the report from Dr. Weissman and others I think he wasn't suggesting that you had to had that evidence.

He was saying that if you have evidence of disparity that meets the threshold at that point you can rest easy and say that's disparity sensitive. And if not then you still figure out what to do.

I was going to say, for looking at
the disparities that are tabulated in the spreadsheet, I think it's important to know that these are absolute differences, I believe. And if you have a measure where it's an adverse event it's rare, then probably a ratio may be more useful.

CO-CHAIR ANDRULIS: Anything else on this question? You want to move on to the next?

MEMBER JACOBS: So my question is should we decide? And it might be helpful, since we're all in the room together, to make the decision now. I hear rousing endorsement for that.

That usually happens with people just sitting around.

MS. MCELVEEN: So how about this? Why don't we let you get some food. We'll take maybe a 30 minute break or an hour?

DR. NISHIMI: Twenty.

MS. MCELVEEN: Twenty, okay, I tried.
DR. NISHIMI: It's Friday, we want to get you out of here.

MS. MCELVEEN: So, again, just a quick break to get some food and come back. And we will continue the discussion.

(Whereupon, the above-mentioned matter went off the record at 12:04 p.m. and resumed at 12:49 p.m.)
DR. NISHMI: Okay, so what Donna so nicely did for the group was to look at the issue of total score versus the percent gap. Donna, can you just take it from here and tee up what's going on.

MEMBER WASHINGTON: Oh, sure, someone raised the question, and I was curious as well, as to whether there was a correlation between the total score, so the score without the quality gap, versus the quality gap.

And what I did was to plot, there were three measures that used units other than percent quality so those aren't reflected here. And there were three with negative scores that I dropped, just to make it look pretty.

And so this just plots the total score versus the quality gap. The total score is on the Y axis. You can see there a bunch of clusters around five and around nine, which
looks like that's where most of the measures scored there, but a scattering of other scores.

And then the gap, which clusters around zero for most of the studies, and that correlates with the distribution of the quality gap but is pretty much stretched out. And just eyeballing it, it looks like there's very little correlation.

For example, if you ignore the three dots in the far right then there's no correlation between score and quality. But thinking about how we might use this data, we might want to consider looking at higher scoring and looking at higher quality gap studies to begin with, for example, just arbitrarily drawing cross hairs somewhere and taking the dots in the upper right corner.

DR. NISHIMI: Any questions or comments? The only comment is, wow, you did that over lunch.

(Laughter)
DR. NISHIMI: She actually did it in much less time than lunch. Liz?

MEMBER JACOBS: So when I saw this, Donna showed it to me, and I was wondering if maybe, instead of thinking about whether there's a gap or not and having a cut-off on quality gap, what we want to do is maybe look at the scores, the high potential scores, and cut it off that way instead of low scores. It's just a different way of thinking about narrowing the field down.

MEMBER WASHINGTON: So high gap plus high score, using whatever cut point we decide.

DR. NISHIBI: And therein is the question, thank you, whatever we decide. So if we can have some discussion on that.

MEMBER EDWARDS: But can I just ask a clarifying question? Are we saying that would be our starting point or that we're actually going to discard some things using that methodology?
DR. NISHIBI: That would be your starting point to look at. But what we would bring back to you then would not be things initially that fell below the line. Or we could bring you back the spreadsheet with all 200 measures at that point, but sorted at the top would be the 30 that were above whatever line you decide.

We're happy to bring you back the full thing but we don't think that's a productive use of your time. That's why we're pressing you to draw a line to sort up. Because then the thought would be that we would have work groups depending on the cross-cut.

So five of you might look at those measures that sort above a line of eight. Another five of you might look at how the measures sort out when you choose a quality gap of five percent, ten percent, 15 percent, whatever you land on, et cetera.

MEMBER EDWARDS: But I guess the
question is, is the goal to eventually get through all of them or potentially not do all of them? Because my only concern, as painful as it might be, with not doing all of them, particularly if there's no data, is that if we don't look at it then there's not necessarily ever going to be any data. And there could be an issue.

DR. NISHIBI: We could create a work group that's assigned to look at all those measures for which there is no gap information, certainly.

MS. MCCLVEEN: So the questions that we present to the group now is, should there be a threshold for quality gap, maybe a suggested scoring approach as you were just mentioning. So something, for example, less than or equal to, or greater than or equal to five, or greater than or equal to ten, just as an example.

And how should we handle measures with high scores but no information on quality
gap? So would these be considered as potential disparity sensitive measures versus definitely calling them disparity sensitive?

MEMBER EDWARDS: So then the two options for that second category would be either potentially sensitive versus definitely sensitive, but not not sensitive? Because I was thinking that, okay.

DR. NISHIBI: Mara?

MEMBER YOUDELMAN: I think I agree with Colette that, on the second question, there is some group that looks at this, makes an analysis one by one as opposed to just saying we're not going to look at these.

In terms of the first question, because I'm still struggling with this, we were talking about it a little bit at lunch, that by saying something is disparity sensitive we recognize that there is a disparity in some of the research available.

But when you get to an individual hospital or provider, it may have a disparity
on that level that is disparity sensitive, it may not. And it may have a disparity on a non-disparity sensitive measure. It can go either way.

So I guess for me, I'd rather have more designated as disparity sensitive than not because we know that there are significant disparities. And by having more, we're putting more emphasis on it. And therefore, hopefully, if someone is looking and using the indication of disparity sensitive measure as their determination whether to look at this or not, there are more things that they can look at.

So I'd rather be over-inclusive than under-inclusive to give more opportunities for folks to think about this and hopefully do something with it in the field.

DR. NISHIMI: Does anyone have any objections to that? But do you have a cut-off point or do you want anything that has
disparities data?

MEMBER YOUDELMAN: This is where you get beyond my knowledge of the research and it's just not my area of expertise. So what is statistically significant, I can't even say the words, I don't know.

So I guess is five lower and therefore more measures are included? So I'd probably say at least five, not 15, not ten. But I don't know if it should go below five to three. That's where I'm not sure I'm qualified to figure that one out.

DR. NISHIMI: Well, at that level it's not a matter of statistically different. It's how inclusive you want to be or not. So if you look at the distribution of the quality gap that we have right now, if you cut it off at less than one you're going to drop 22 measures. If you cut it off at five percent it's 20 more measures.

MEMBER YOUDELMAN: Then this gets back to the discussion earlier of, if we're
too inclusive and there's not a significant gap, have we pushed too far. And people are going to push back.

So that's why I'm having a hard time. I certainly think five. I would probably argue for lower than that but would want to get other folks' input on what they think is appropriate in figuring this out.

DR. NISHIMI: Kevin, I'm going to put you on the spot because you were on the ambulatory sensitive measures when we did that project. So you've been with it from the beginning.

Do you have any thoughts about, because we didn't do this kind of ranking there. It was more, you looked at it. Do you have any thoughts on where we might draw the line?

MEMBER FISCELLA: Yes, I don't know that there's a clear answer. I think there's two competing issues here, at least that I see. One is the issue of focus and saying,
we're going to focus, perhaps as a country, on these as high priority areas and really try to hit them and do them well, versus the competing need for inclusiveness and lots and lots of measures with the potential for less focus and less movement.

So I don't think that there's a magic answer here. But I think it is important for the group to think about how many measures we want to have at the end of the day and to be thinking about those two competing issues, the issues of inclusiveness and representativeness versus the opportunity to focus and perhaps make a greater impact on fewer.

DR. NISHIMI: I'm sorry, Marshall, then Ellen, then Ernie.

MEMBER CHIN: It's a question maybe for Ernie and others who may know. So I'm assuming that these quality gaps, I don't know if it's correct, are these national numbers? It really depends upon the measure
in terms of where the literature is coming from.

In other words, the question for Ernie really is well, what do we know in terms of regional variation, such that even if a measure maybe, aggregate from these studies, have zero disparity.

If there's significant regional variation then it may be something that we might still consider in terms of giving regions or organizations the flexibility to pick things that are relevant for them.

MEMBER MOY: Yes, you can pick anything and there'll be so much regional variation that you're going to find disparity. So that's just the way it is.

DR. BURSTIN: And just to follow-up on that most of our measures are used at a national level. So I think we're trying to keep it applicable at this level. But I also think communities could use these criteria to help understand, within their
community, which measures they would always want to stratify on.

So I guess the question is it also just a useful sorting tool for a community to help think through where they may have issues and they should always stratify.

DR. NISHIMI: Ernie, you had your email, and I'm sorry.

MEMBER MOY: Yes, instead of this being a flagging thing, disparity sensitive or not, can it be a label that quantifies the amount of disparity, that it's been observed in the world?

So there might be a category for large disparities demonstrated for something that's more in ten percent, and moderate for five to ten percent, and small disparity for zero to five percent. Does it have to be a yes, no kind of variable or can it be something that is more descriptive in nature?

DR. NISHIMI: I don't see why not, do you?
DR. BURSTIN: Marshall, do you want to speak to what went on at the MAP because I think that has some direct relevance in terms of selection of measures? Do you want me to do that, you keep looking confused.

MEMBER CHIN: I'm not sure if I know what you're talking about, Helen.

DR. BURSTIN: Marshall, I thought it was your suggestion. In a parallel part, I think you asked the part about the measures application partnership. They're helping with the selection of measures for pre-rule making on the part of CMS.

So one of the criteria we actually put into place for how they would look at an overall set of measures for a given program, like the in-patient quality rule, or the out-patient rule, or the nursing home rule, or the home health rule, is one of the criteria.

We said, are there just disparity sensitive measures building prospectively what we hope this will provide. So the idea would
also be that that flag is important.

Although underneath it you should be able to see the data. Because I think we'd like to make sure that in all of these programs they are, in fact, pulling in some measures where there are disparities and they should be looking at, and asking hospitals and others to stratify.

So it actually has a direct applicability to the selection of measures. So I thought that was your suggestion Marshall, sir.

DR. NISHIMI: Ellen?

MEMBER WU: I guess my question is, in your previous experience about putting out measures, has there been a number of measures that seems doable, not overwhelming but enough that it's comprehensive, a range?

DR. BURSTIN: In the ambulatory care project we selected what, about 35, I think, Kevin, ambulatory care sensitive measures out of a couple of hundred, is that
about right?

So I don't know that we know what the right number is. But I think, as you've all pointed out, it's a balancing act between wanting to keep focus on what's most important and yet not wanting things, where there's potential disparities, to not get looked at so that you could find where there are disparities. So I think it's really a judgement call on your part.

DR. NISHIMI: Jerry and then --

MEMBER JOHNSON: Yes, in the interest of us making a decision at some point, let me make a concrete suggestion that just builds on what others have said.

First of all it sounds as if we've said that maybe eventually we were going to try to do almost all of these anyway. So the real question is how we prioritize what we do first. It seems like that's what we're deciding.

And so suppose we say that we
would start with those with a quality gap over five and that would give us about 28. And then if we go back to that chart that Donna had before and look at just the total score and take total scores above some number, actually say ten or more, then that would pick up some additional ones.

And then what I would do is look at that group and see, in the categories here, whether we actually at least have a measure under each type of condition.

So even though there are 20 measures or 20 conditions, there are not 20 different types of conditions. There are four or five here that are cardiovascular and a few that are cancer and a couple that are behavioral health, and so forth.

So I think we would want to have at least one measure in each of the different types of conditions. So I would just do it in that stage.

First, quality gap cut off, then
total score, and then get some measure from every condition and say this is the first set that we'll look at.

    We'll look at this set first, maybe now we have 35, 40 measures. I don't know what it would be. And let's do those first and then move on.

    DR. NISHIMI: I think that's good.

    Are people comfortable with that? And then what we could also do along that line, Jerry, thank you for that suggestion, is in addition to the conditions look at the settings.

    So that if they all end up being hospital measures then we'll pull up nursing home, the SNF measures, we'll pull up the top two, home health, and do it that way. Okay, I think that's good guidance for the taggers, as we call them.

    DR. BURSTIN: It also has to be interesting as we're going through the next set of projects. We're doing pulmonary right now where just the committee meeting's coming
up next month.

It might be interesting to actually pull some of those data from those just to give you that flavor of a prospective set, of how you might do this going forward with the data that comes in.

DR. NISHIMI: And so with that in mind I would like you to at least take away the second question, how are we going to handle measures with high scores or medium scores, whatever. But we have no information and we heard conflicting views on that.

Not something that we have to decide today, but I do think that the committee's going to have to land on a justification of anything that it might move into potentially disparity sensitive. Or I would argue you couldn't classify it disparity sensitive, personally. But some of you might want to.

So you're going to have to do some thinking around that. And it would be useful
if you could start thinking about it now while
the issue's fresh in your mind and email
Nicole your thoughts on how we're going to
handle that.

Even if you call them potentially
disparity sensitive, or high, medium and low,
which I liked Ernie's suggestion, you're going
to have to justify how you got to those
places. Helen, you have any, oh, Kevin, I'm
sorry.

MEMBER FISCELLA: I was going to
say, related to that issue is what are next
steps after the committee? Because I think it
would be easier for me if I knew what the plan
was in terms of data collection for all of
these areas where we don't have any data.

And we have no idea, or don't have
a good idea, of whether there are disparities
there and who might be affected and how big
they are.

DR. BURSTIN: It's a great
question and thinking more about the
prospective arm of this, we talked a little
bit this at one of the in-person meetings
about our submission form.

But we will be re-doing our
submission form this summer with a pretty
large scale overhaul. We're actually going to
be moving, we think, to splitting out the
endorsement process into two stages.

So that the first stage will be a
review of a measure concept, really looking at
importance, evidence, a lot of the issues you
guys really tangled with yesterday.

And then if you pass stage one you
get to stage two where we'll look at the fully
tested, fully specified measure. So the idea
is a lot of people invest a lot of money and
resources in developing measures that never
get past importance because the evidence isn't
there. There isn't a quality gap, et cetera.

So we're going to be doing
significant work on the submission form this
summer, in short, to be able to split it out
and think through that process. So maybe one idea would be to actually have you take a look at the questions, perhaps in more detail than I think we did last time.

And just say, in light of this conversation, what prospectively would you want measure submitters to submit, at either the concept stage and the fully specified tested measures stage, that would allow you to automatically come up with an algorithm that says, yes, this measure should be classified as disparity sensitive and prospectively stratified.

Not that they have to answer today but we're happy to engage in that. I would find that incredibly useful because I think, as much as it's wonderful to bring this to you, we want to make this part and parcel of the work of NQF.

So that every kind of measure comes in, there is an assessment of that. And it does sometimes depend on how, Ray Gibbons,
who is the chair of cardiovascular, who is the Chief of Cardiology at Mayo, President of the American Heart Association, could not have been more strident.

Any measure without disparities data up for maintenance was sent back. And they needed to run it, get it back, or he wouldn't look at it. And it was great. So we're not trying to stick with that.

If your measure's been out there for at least three years and you've got nothing on how it's being used or what the disparities are, well, you can find it. Bring it back when it's ready.

And so part of that other process is we'll move to almost a batch production line for all of these areas. So we'll allow measures to be submitted like every six months across all these areas. So you don't have it, go back out, finish it, bring it back in six months.

So we'll have a lot more latitude
to have people just go away and get the data, fix it, bring it back when it's ready. But your input would be great there --

DR. NISHIMI: Does that help you, Kevin?

MEMBER FISCELLA: Yes, it does. I think it's really important because one of the problems is that, I think, when many people look for disparities and they don't find them they may not publish it.

So a lot of this data is never published. So then you don't know. Did anybody look and find it or was it not published?

But by asking people who are coming in with new measures to begin collecting that data and presenting it, I think creates a much richer environment to really assess where the disparities really are. And helps me to feel more comfortable about moving ahead, at least in the areas where we know there are disparities and then
filling in as we go.

DR. NISHIMI: And the hope would be that we would be through the portfolio in, what, two more years. So in the greater scheme of things that, for our processes, a pretty short time frame. Francis?

MEMBER LU: Yes just a question. In terms of question number two just from the NQF protocol point of view, to what extent would it be possible for those measures that don't have any disparities data right now to group them together, or maybe a sub-set of those depending on what we decide, and label them as potentially disparity sensitive based on certain criteria, worthy of further investigation or something.

Is that possible? Because I think that still would be beneficial to the field because this is such a new field. And the work that's been done so far, that has provided some assessment of these measures, would really jump-start the whole process so
people don't have to start from scratch in devising measures and all of that business.

But here's a set of things that have reached a certain quality level that bear further investigation. If that fits the NQF protocol, I think that would be very helpful.

DR. NISHIMI: Yes, Helen's indicating that. And certainly what it would do is alert the measure stewards of those particular measures, that there is an expectation, when you come in to have your measure re-reviewed for maintenance, you really need to come in with the data. Okay, any other thoughts on the assessment? Nicole?

MS. MCELVEEN: So next step, so immediately the first thing that we will be doing is reviewing the RAND measure, as I mentioned earlier.

And what I would like to do is provide the materials for that measure to the group on Monday. And that's this submission form and then the full survey itself for the
group to start to look through. And you will have about a week to look through that.

And we want to just get a conference call scheduled pretty quickly to try and get that completed. The other conference call that we definitely will have is one that happens after we complete our comment period.

So once the report goes out for comment and we get those comments back we review a good portion of those with the committee to get your feedback on how we should respond to certain comments.

Many of those we do defer to the measure developer to answer because they're usually the ones who know the response to those questions. And there probably will be instances that we'll need to get feedback from the group.

And the measure developer will be making changes only to that one measure, again the measure addressing qualified interpreters.
That change will be made to that. And I think that's it on the process side of next steps.

The other piece I wanted to go through with the group is around our timeline for moving forward. Our comment period is scheduled. It's a 30 day period. It's scheduled to open in May.

The dates listed, Adeela, sorry, you guys have no clue what I'm talking about. Let's see, there we go. So looking at our comment period, again, is going to be from April to May. Conference call in, looks like the third week of May to review those comments.

What then happens next is we go out for an NQF member vote. That's a 15 day period for members to vote on the measures. We traditionally have a pre-voting webinar.

That's just an opportunity to reach out to all stakeholders and all groups to let them know that this report is coming
out for vote, to answer any questions that may arise before that vote period happens.

And then later this summertime, and then into August, is when we'll conclude the project. And that will happen with a CSAC decision. The CSAC will review our set of measures and endorse the measures that they feel are appropriate. And the board ratifies that decision and then we have an appeals period in August.

So the important thing that will happen immediately following this meeting will be the RAND measure will be circulated to you. I will also circulate a survey that will poll you for availability for conference call dates.

We then will sort out the measures assessment piece of the project and figure out the best time to meet with the group to go through the final steps of those. And, Grace, did you have a question?

MEMBER TING: I did, so for the
gap measures that we identified earlier this morning, is that just going to be future iterations, we'll just leave those as gaps for this round?

MS. MCELVEEN: Right, so that information will be included in our report. Did you want to say something about that, Helen?

DR. BURSTIN: Actually Dennis made a suggestion before he left that, just given the brain trust here. It might be really useful, perhaps, to have us send out that list of measure gaps to you, that you all came up with, and actually have people even sketch them out a tiny bit more in terms of more of a measure concept, so that we actually provide a bit more information to the field in terms of where, a bit more specificity to measure gaps.

MS. MCELVEEN: And the other thing that I wanted to remind the group is pertaining to the measures. If you have any recommendations, particularly around the
measures that were not endorsed or that were not recommended for endorsement, recommendations for the developers, I know, Colette, you had asked about that earlier, just to please email me that information so I can pass it on to them. Are there any more questions? Kevin, did you have a question?

MEMBER FISCELLA: Yes, I just wanted to get a sense of whether you feel there's a clear enough consensus of where to go here, given that this is going to be, I guess, our last in-person meeting. Is that right? And Liz's early comments on whether we need to be clear or whether this is clear enough.

DR. BURSTIN: I think we'll have a better sense of that when we digest this. We also could potentially, if you think it's important, try to actually add another meeting. But let's just see, in the post-meeting analysis, how clear Robyn and Nicole are feeling.
DR. NISHIMI: I do think that, in the short term, the guidance that we got today around the measures assessment was absolutely excellent. And so it's clear the next steps that can be handled.

Whether or not, once we get through that and we see that we now have 300 measures that are in the no data category, what to do about that.

Whether or not, as Helen said, we have to go back and re-think our strategy, do a couple conference calls with you that don't prove to be fruitful and meet again, I think. To me that is the biggest issue right now.

I think once we do those other cross-cuts and sorts you'll be able to work through those. And that's not the issue. It's not entirely clear to me what we're going to do with those other ones.

MEMBER JOHNSON: I just had a quick question. You might have said this earlier and I missed it. What happens with that
information about the gaps we identified?

Do you go out and search other members, do you put out more of a directed call for those things, or what happens with those gaps we identified?

DR. BURSTIN: So they'll certainly be in the report, as we were just talking about. We can, again, you guys are connected to a lot of the organizations where that information might be useful. You should feel free to distribute it if you think there are developers thinking in this area of what to work on next.

We can certainly work through Ernie and others to see if there's some opportunity there. But we do routinely have measure developer webinars we conduct every month. About 80 different developers come on on a monthly basis.

So we do routinely try to, it may be a very good opportunity, maybe on one of these upcoming calls, to describe to them what
we're going to be asking to them, very clearly up front, as they submit their maintenance measures. And here are the gaps that were clearly identified.

We've been trying to encourage them not to spend a lot of energy on look-alike measures as we like to call them, like look-alike drugs, like same old measures, different settings, same old measure, a different slice of the population. But actually, hopefully, invest those very limited measures, all the dollars, where we need them, like these gaps.

MEMBER CUELLAR: I'm sorry, Elizabeth just triggered my memory. I actually identified one other gap if you don't mind, and that is persons with disabilities or functional limitations.

The ability just to get a PCP, or just general internal medicine type of care, dental care, it's just the accessibility issues are very, very wide.
And the other thing would be transitional care for children with disabilities transitioning to adult care. We have children being seen by pediatricians that are in their late 20s because they're not enough physicians who will take children with disabilities. It's an issue of accessibility, a lot of issues surrounding that area.

DR. NISHIMI: Grace and Kevin, Kevin?

MEMBER FISCELLA: In some cases there are conditions like sickle cell anemia, and management of pain in sickle cell anemia, where it really almost exclusively affects one group with a common ancestry.

And that wouldn't be a true disparity, but it may be a disparity in terms of the fact that this group has pain managed less optimally than other groups who experience pain. And so that's one potential gap.

Another potential gap is the
correctional population, which gets very little attention. And particularly in healthcare that relates to treatment of chemical dependency as well as mental health.

But the medical side is probably not as bad but in terms of mental health and substance abuse it's just abysmal. And of course that's what gets people incarcerated to begin with, oftentimes a behavioral problem.

But yet the care within this group, that disproportionately affects poor and minority people, there's very little oversight, very little reporting and very little public accountability.

DR. NISHIMI:  Go ahead, Mary.

MEMBER MARYLAND:  So I would echo the issue around pain, not just in terms of sickle cell anemia patients, it's a matter of chronic pain, and as importantly, chronic pain control at end of life. So those are really big areas where frequently most providers and institutions don't nearly adequately treat.
DR. NISHIMI: Anything else, Nicole?

MS. MCELVEEN: Nothing else from my end. I do really want to thank the group again for being available, being attentive, and helping us get through this information over the last two days.

DR. BURSTIN: And again if you have thoughts, big picture thoughts, of what you think, in our role here, we could help with in this field please let us know, thanks.

DR. NISHIMI: Thanks very much, everyone, safe travels to those of you traveling.

(Whereupon, the above-entitled matter went off the record at 1:24 p.m.)