THE NATIONAL QUALITY FORUM

+ + + + +

PATIENT OUTCOMES STEERING COMMITTEE

+ + + + +

MEETING

+ + + + +

Tuesday, October 20, 2009

+ + + + +

The meeting convened at 9:00 a.m. in Salon D in the Marriott Metro Center, 775 12th Street, N.W., Washington, D.C., Joyce Dubow

and Lee Fleisher, Co-Chairs, presiding

MEMBERS PRESENT:

JOYCE DUBOW, MUP, Co-Chair
LEE FLEISHER, MD, C0-Chair*
RUBEN AMARASINGHAM, MD, MBA
LAWRENCE BECKER*

E. PATCHEN DELLINGER, MD*
ANNE DEUTSCH, PhD, RN
BRIAN FILLIPO, MD, MMM, FACP
LINDA GERBIG, RN, MSPH
EDWARD F. GIBBONS, MD
LINDA GROAH, RN, MSN, CNOR, FAAN
PATRICIA HAUGEN

DAVID HERMAN, MD*
DAVID S.P. HOPKINS, MS, PhD
DIANNE JEWELL, PT, DPT, PhD
DAVID A. JOHNSON, MD, FACP, FACG, FASGE*
IVER JUSTER, MD
BURKE KEALEY, MD, FHM
PAULINE McNULTY, PhD

MEMBERS PRESENT (Continued):

VANITA PINDOLIA, PharmD, BCPS*
AMY K. ROSEN, PhD*
BARBARA J. TURNER, MD, MSED, MA, FACP*

BARBARA YAWN, MD, Msc, MPH, FAAFP

STAFF PRESENT:

HELEN BURSTIN
SARAH CALLAHAN

JENSEN CHIU
ALEXIS FORMAN
MELISSA MARINELARENA

EMMA NOCHOMOVITZ

KAREN PACE

REVA WINKLER

BONNIE ZELL

MEMBERS NOT PRESENT:

SHELDON GREENFIELD, M.D.

*Via Telephone

TABLE OF CONTENTS

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

(9:02 a.m.)

CO-CHAIR DUBOW: Good morning again. Hope you all had a nice evening last night.

Day two, we probably will arrange our agenda just a little bit in order to be able to break early, depending on how our conversation goes.

We're going to talk about the measure evaluation criteria. And again, we have Karen Pace to walk us through some of that material. Helen will join us later.

We also have Linda Gerbig, who joins us, so, Linda, we'd like to welcome you and give you an opportunity to just share with us who you are. Go ahead.

MS. GERBIG: Thanks. I'm glad to have finally arrived and to not be on the telephone today. That's an absolutely miserable experience.

I'm an RN by trade, a nurse by

trade, and I'm the vice president of performance improvement for Texas Health Resources, and we are a 14-hospital system across North, Central and Western Texas, not-for-profit, faith-based organization.

MS. GERBIG: Great. Well, welcome.

Pat, you had a really truncated opportunity yesterday. Maybe you'd like to re-introduce yourself, too.

MS. HAUGEN: Okay. I'm Pat Haugen. I'm an inflammatory breast cancer survivor, 12 years, and I do volunteer work with the National Breast Cancer Coalition.

My business career was with IBM, so I have some experience with quality not specific to health care measures, but have been on one NQF panel for physician-level oncology measures. And then our organization has done some work to train advocates in some of the specifics on measures relative to cancer, and breast cancer, specifically.

Thanks.

CO-CHAIR DUBOW: Thanks very much.

Okay. If there are no other issues at hand, let's get started.

DR. WINKLER: Okay. I just wanted to make a couple comments, and thank you all for your discussions yesterday.

I spent some time kind of reviewing and figuring out what it was we talked about. Lots of food for thought. And I would like to encourage all of you who keep coming up and whispering in my ear and showing me some of the things that you're doing and ideas and thoughts.

Keep the cards and letters coming, because some great ideas have really stimulated my thinking on terms of how to organize some of this stuff. So, I really do value and need your input. That's how the steering committee can work very effectively in shaping the way this project goes forward. So, thank you very, very much.

Just as a follow-up, we had a lot of talk about principles and definitions yesterday, but I don't think we kind of came to any conclusions, and so one of the things we will be doing is, from all the discussion, the notes, the recordings, all of it, is drafting what we think you said or think you wanted to say or tried to say, or something, and then send it back out for you to then, you know, work with it and see if you can come to do the wordsmithing, be sure the meanings are correct so that we can kind of capture all the good thinking that was there.

Though I still think it's still in rough form, I think we've got the kernels from which we can get some good final product.

Okay. This morning's conversation is about some of the nuts and bolts of NQF work. For the part of the project where we're going to be evaluating candidate measures for endorsement, it's really critical that the steering committee understands the whole

process of evaluation and NQF's evaluation criteria.

As I mentioned briefly yesterday, the whole process and the whole criteria over NQF's lifespan has evolved, it's matured, it's become more rigorous, and it's become more focused with some very specific reasons for that.

And we are trying to reach a higher level of performance, recognizing that performance measures are tools, that can drive performance through a variety of mechanisms. And so, we want to keep pushing it, raise the bar higher, pushing harder.

So, as I mentioned yesterday, the evaluation criteria were revised a year ago in an attempt to meet these sort of higher expectations.

And a lot of our processes are also evolving. We're trying to move into a fairly exclusively electronic world. For those of you who have worked with this before,

you realize that a lot of trees were sacrificed in our behalf, and so we're trying to, you know, keep everything electronic.

In doing so, a couple of things around the evaluation and the measure submission process are now as automated and, hopefully, are going to become more and more so as we go forward.

But probably one of the most significant changes for us was, this summer we were able to institute an electronic measure submission form.

Previously it was a fill-in-the-blank Word document that, you know, then we had paper. Lucky us. And it's a fairly voluminous amount of information to manage.

Now that we've got it in electronic form, we can then reformat it and change it and give it back to you in any old way we want.

So, this has been a change. All changes do not necessarily go totally

smoothly, and they are not without their own individual issues.

My friend Karen, who will talk a lot more about the evaluation criteria has been very much involved in working out the bugs, working with the contractors, kind of make this process work, because it's still in evolution, and trying to make it work for today, when they keep saying, "Well, in six months we'll be able to do this."

It's like "Great! But what are we doing today?" So, Karen is our how-are-we-going-to-do-it-today kind of person.

So, the measures do come to us now in an electronic submission format. For those of you who have contacts with measure submitters or are, you know, potentially someone who will submit measures, the information on measure submission is available on the website, and you can, from our project page, because we have an open call for measures right now, go to the measures

submission form.

Go to the Call for Measures section, open it up, submit measures. There's a button. You actually need to have registered for the website in order to submit because then whoever is submitting has their profile.

Dashboard. That's what we call it. Dashboard. The wrong word, to follow what happened to that submission.

So, if you're a measure submitter, there's a reason to be registered and all that kind of stuff. So, things are happening in that realm.

So, luckily now, as opposed to a bunch of Word documents, we now have spreadsheets with a whole bunch of stuff in them. So, that information is now available to us electronically.

Thanks, again, to Karen's good work, the output that we're going to be giving you is something we've given you an example

of, and we're going to talk about today, because we asked the questions in ways that we hope work for the measure submitter.

But for the evaluation, we want to use that information, aligned with the measure evaluation criteria, so we reshuffle it, and put it in a different format, something that was almost impossible because it was so cumbersome in a paper world.

So, again, Karen has worked with that. So, we want to go through that with you.

Part of the evaluation process, though, does have a lot of up-front work, and so staff has work to do. The steering committee has parts of it, the TAP have parts of it, and we're kind of building an evaluation process that gets more robust over time.

So, this sort of outlines the process. The staff, our staff will be evaluating whether the conditions for

consideration are met, prepare and distribute things. So, we're the paper-pushers, if you will, the electronic paper-pushers. Thank you very much.

And this steering committee will be primarily reviewing the cross-cutting measures. There isn't a TAP for the cross-cutting measures, all right, so you won't have that step.

The TAPs are going to evaluate the subcriteria, and if that's a confusing term, hang in there, we'll show you what we mean for the condition-specific measures appropriate to the different areas.

Then, the full steering committee will evaluate and vote on the threshold criterion, which is importance to measure and report, and we're going to talk about that.

And for measures that pass that criteria, then we evaluate the remaining three major criteria.

The full steering committee, then,

votes on those recommendations regarding whether to recommend to the membership the measures go forward for endorsement.

So, that's sort of the outline of the process, but we want to take you through the evaluation criteria, because our experience is such that sometimes it's hard for steering committee members to grasp what we're meaning, why we've got the criteria the way we do.

And this is an opportunity to kind of go through it in detail with you and you can ask questions and we can all try and get onto the same page.

As both Joyce and David will tell you, the more the steering committee can work within this construct, the cleaner things come down the rest of the steps, and we don't have to do a lot of "Send it back," or "What in the world were they thinking?" or "Why did they did they do that?"

And it facilitates the

communication if we're all working off the same sort of rules of the road, if you will.

So, the evaluation criteria that we use is standardized. There are four main criteria. The first one is importance to measure and report, and we very specifically have stated it that way. In the old world it was importance.

But important has so many meanings to so many different people. We are all about these performance measures being important to drive quality improvement. That's what we're here for. All right. There are so many very important things out there, but not all measures that are developed have the characteristics and the capability of doing good things as a result of implementing them.

So, it's an important threshold criteria. We're going to talk more about the details.

Scientific acceptability of the measure, itself, as opposed to the science of

the topic that is being addressed. It's not just are beta blockers good in patients with coronary artery disease, is this measure, the way it's specified, you know, precisely specified, reliable, valid and all the other good things necessary. So, we are really in that category, looking at the measure.

Usability, it's the "So, what?" question. Okay. Somebody does it, collects the data, has the data, can they use it? Can they use it? Do they understand what it means? Is it useful for a wide variety of things, particularly for public reporting?

And then feasibility. Can it be done? Is it even possible? Great idea, but can it be put into production such that it can be used in a widespread way.

So, we're going to talk about the subcriteria that helped feed into those, because again, so many of the questions that come back during comment period, that you may get asked back by the CSAC or even potentially

the Board of Directors, generally you're embedded in one of these somewhere and without appropriate consideration of them, you end up kind of wishing you'd spent a little more time thinking about it.

So, I'm going to ask my friend Karen to jump in at any time. Karen was the staff person who worked with the subcommittee of the CSAC to revise the criteria last year. So, she spent endless hours of these conversations of how the criteria should be characterized to try and reach the goals that we've set for what the endorsed measures should do.

So, importance to measure and report, this is looking at the specific focus of what is measured, and it needs to be important enough to expend the resources to collect data, analyze the data and report the data.

All right. So we're talking about a balance. There are lots of very important

things, and that's why we try not to use the term "important," because, you know, that's a value system, and what's important to you is important, but we're talking about the importance of this measure in measurement reporting within the NQF world.

And so, it's a balance, because measurement is not free. It's costly. So, we need a bang for our buck, and so that's the importance to measurement report. So, not that it's important in its own right, but that the measure, the actual measure, has importance.

And this is one area that was really worked on very significantly in the revisions. There are three. One is relationship to an NQF priority, and if it's one of the NQF priority or the priority partnership's goals, aces, and we actually have a special section on the form, and then we'll check it and tell you, flag it and say, "Hey, this one of them," or, because not

everything falls under all those priorities, a high-impact aspect of health care.

Now, impact can be a large number of things, number of people, dollars spent, severity of illness, you know, those sorts of things.

So, impact, again, is in the eye of the beholder, but it is important that it has some oomph behind it in some way, shape or form.

Importance, and you may think this is a little bit inconsistent, because so many people say why isn't this in the science, but the evidence to support the measure focus. Okay. If their process measures were looking for the relationship to outcome, what is there, what is the evidence, how good is the evidence, what is the evidence that says doing this will get you what you want.

And so good ideas might be very important, but without the evidence behind them that really gives us a strong tie to good

patient outcomes, you know, maybe not so good.

In this particular case, we kind of jump this one because all of our measures should be outcome measures. And so, outcome measures sort of in and of themselves, reach that higher level of criteria of being outcome measures.

However, at the same time, perhaps not all outcomes, which will be your realm to determine, are the most important things going.

Opportunity for improvement. Again, some people will short-cut this to say the gap in care, but it's not just that. The question is, at the end of the day: do you envision that if this measure is put into play, and measurement we do know changes behavior, that we will see improvements in overall health care, in overall outcomes.

So, the opportunities for improvement may be current lower performance, but it could also be variation. So, maybe

you've got some folks doing really well, and their mortality rates or their complication rates or their intermediate outcomes are just stellar, but there are a whole bunch of folks that aren't doing so hot.

So, you know, we've got this variation. We want everybody to experience good care. We want to raise all the boats. So, the opportunity for improvement in variation in care is also an important aspect.

Also, the opportunity for improvement, as Helen alluded to, may be a combination of the impact. You get a lot of people, even if you're only going to move it a little bit, you're going to move a little bit over a lot of people, and that may ultimately have a significant improvement.

So, part of the challenge to the steering committee is weighing these in determining if it's important to measure and report and move it on further through the process.

The key aspect of all of this is, if you decide it's important, why? On what basis, so we can say it's because it's this, it's because it's this, it's because it's this.

Conversely, if you say it's not important, then it's not important because it doesn't meet these criteria.

This is the way we communicate to all the stakeholders out there, because, as you know, in a multi-stakeholder world, there are folks who are tied to certain things, and some measurements are going to be very important to them, but perhaps not in a greater world.

So, we need to be able to explain the decisions and not just say, well, I just thought so. That one's hard to justify and hard to move forward.

Yes. Karen, jump in.

DR. PACE: Just a couple other things to building on what Reva's already

mentioned, and that is with the NPP, you know, we really are looking for the specific goal, but as Reva said, if it doesn't address an NPP goal, that doesn't mean the measure is out. You know, there's all kinds of way to look at high-impact. So, that's not a reason.

The national priority partners.

DR. WINKLER: That we talked about.

DR. PACE: Thank you.

PARTICIPANT: Six priorities we mentioned.

DR. PACE: Right. And there are some specific goals attached to each of those priority areas. So, that's what staff will be looking for and provide that information to you.

The other thing is that all of the things that Reva talked about, about, you know, the opportunity for improvement and the evidence, we really are asking the measure submitters to provide some data.

So, rather than just saying there's variability in performance, if there's any studies that have been done in the literature, even if it's from some pilot work that they did, we're trying to ask people to provide some context for saying that it's a performance gap so that you have something to look at.

And, as Reva mentioned about the evidence, we have gone back and forth of where we situate that, so it's interesting, but the idea is that, and some of our earlier documents talked about leverage, and Reva was talking a lot about, you know, are we measuring things that are really going to move us forward in improving health care.

And so, this idea of leverages, if you're doing things that have really been proven to improve outcomes, that's what's going to, you know, warrant measurement so that we continue to make some improvement in those areas.

And because it's a threshold criterion now, we thought it best to be included in that importance criterion.

DR. WINKLER: Yes, that's what I was going to say. Just to reiterate, the must-pass, and it kind of stops things right there. So, just to reemphasize.

Okay. Now, questions.

DR. JEWELL: So, I think one of the interesting things about the evidence piece that you just discussed, is that many of the outcome measures, at least in our world, were designed, as we talked about yesterday, at the patient level, and really were not designed with quality improvement in mind.

DR. WINKLER: Right.

DR. JEWELL: So, the availability of evidence, I think, is going to be pretty variable, depending on which outcome measures we're talking about. And I would venture to guess, probably pretty scarce initially, just because they weren't designed as provider-

level measures in the first place.

DR. PACE: And as Reva said, you know, for all of our other types of measures, we're looking for evidence of association with the outcome. So, when we're starting with the outcome, it's the question: What would the evidence be?

We have asked in this round of measure submissions for submitters, if they have knowledge of studies that have shown that that outcome can be improved. It's not an absolute requirement because the way we look at outcomes for quality improvement is the variability that Reva was talking about.

If there are some providers that are doing really well, it shows that you can't, you know. And if it's a proper measure, risk-adjusted, it shows that you can achieve higher levels, but it is helpful and strengthen things if there have been studies done that show that improvements can be made in that area.

So, we ask for that, but it's not like something that would necessarily stop a good outcome measure from going forward.

DR. WINKLER: Yes. This is really the difference between evaluating process versus outcome measures, one significant thing is where it's a very critical part of evaluating a process measure, how do you determine the link to outcomes on an outcome measure? I mean, you know, it sort of negates itself, if you will.

Are there any other questions, because this actually is the criterion that steering committees wrestle with the hardest, because they want to keep that "important" concept, you know, tightly bound. Well, this is really, really important. Yes. Okay. Good.

But that difference is sometimes difficult, and we appreciate that, but that's why we're trying to get you indoctrinated, if you will, into kind of thinking of importance

the way NQF thinks of importance.

DR. YAWN: Well, I think that's one of the potential advantages of having sort of multi-perspectives, because you guys are very tied in to rehab. I have to tell you that most of my patients never go to rehab. They don't need to go to rehab.

So, I'm going to have a very different perspective. When you tell me, "Oh, this is the most important thing in the world," I say, "Yes, it is for the eight percent of patients that do it, but it's not too important for the 92 percent that don't need it."

So, I think that will be one thing that's helpful, and I look forward to lots of people sort of pulling my chain back and saying, "No, no, you're not looking broadly enough."

DR. GIBBONS: Would you say that, then, in this context, every measure has to be an outcome measure and a process measure has

to be associated with an outcome measure?

DR. PACE: That's what we would prefer.

DR. GIBBONS: Right.

DR. PACE: But we know in reality there are thousands of health care processes that haven't been studied with the evidence that we're talking about, but the ideal is if it's something, you know, important enough to drive changes and improvements in patient outcomes, that there's some evidence behind it.

DR. WINKLER: I'm going to take it one step further, because not only do these measures have the potential to drive improvement, they are used for accountability.

And if you're holding people accountable in any variety of ways, you really want something that's based on some pretty strong evidence as opposed to, well, it seems like a good idea, you know.

And that is one of the reasons we

want to keep the threshold fairly high, is the impact on ultimate uses for accountability.

CO-CHAIR DUBOW: But in terms of linking it, you know, having a known relationship to an outcome, I think it's fair to say that among all of the measures in the NQF portfolio, there is quite a degree of variation.

And the point is that, you know, we're trying to raise the bar, as we said yesterday. So, as we keep raising the bar, I think the likelihood of seeing more measures that have that known relationship to proliferate and see the other ones go by the board.

DR. WINKLER: And a lot of it is a relative thing. We will often see a lot of measures based on, say, guidelines, but the level of evidence is Level C consensus. There's no science behind it.

So, you know, is that evidence or is it not evidence, you know? And so, it

isn't as black-and-white, and there are gradations, but luckily for the outcomes project that is less of a concern for us.

But you can see why it becomes a real significant issue for the steering committees.

MS. GERBIG: Yes, and another issue that I just raised is, without evidence from the user's point of view, the user of these measures.

We spend all of our time storming and arguing about the validity of the measure, and quite frankly, we never get on to improvement because we can delay it by arguing it.

And I think one of the reasons also, as users, we tend to sort of glom onto a process measure, as you can measure them and report on them quickly. So you can do something very quickly.

But I am not aware of an outcome measure that you can get with any sort of real

time data that's actionable, so then you storm into the issue of, well, this data was two years old, or more than a year old, so I fixed that, and I don't have to pay any attention to it, only to wait another year to see that it was never fixed, and you're back into the same position again.

So, I think that's something we're going to have to deal with on the outcome measure issues, along with the evidence.

DR. HOPKINS: It seems to me, if we do our job here, we will have defined the outcomes and identified the measures of outcome that are what we want process measures to be linked to. No?

But this is the outcome steering committee.

DR. WINKLER: Okay.

DR. JUSTER: I had one question of Linda. Maybe I didn't understand where you were going with that it can take a year or two to get the outcome. It certainly would be if

you were looking at, for example, hospitalization rates for something you wouldn't want to track changes in that every week, necessarily.

But would some outcome measures such as presenteeism, possibly some kinds of functional status, and then the intermediate outcomes, of course, like blood pressure or something, naturally you would track.

I'm looking at the other side of importance, the accountability that you were talking about in terms of driving the outcomes improvement system directly from the outcomes measurement system.

If I was building a tool inside of a, you know, an office, I mean, a facility, I might want to have the outcomes measurement system directly feed into the outcomes improvement system.

MS. GERBIG: And, you know, we have the opportunity, maybe, to begin to think about outcomes in a different manner than we

have historically considered them.

For instance, all-cause mortality or all-cause readmission on hospital comparers, what we typically think of as an in-patient outcome measure now, but does it need to be.

I don't know. Are there outcome measures that we could measure much more real time than we do and we have sort of taken the easy way out in the way that we do it now.

So, I would agree that there could be some possibilities that we've just not looked at in the past.

DR. WINKLER: Yes. Linda, one of the conversations the committee had yesterday was on the types of outcome measures and I don't want to scroll all the way back through that slide, but there were any number and the group added a few more.

Certainly mortality, certainly complication rates, certainly service utilization, like readmission, however, there

were other things such as the intermediate outcomes, things like functional status, things like patient-reported outcomes around symptoms, or how are you feeling.

So, we were discussing that wide spectrum. So, I think to the degree that there might be measures in existence out there, we would certainly want to be able to look at it from that perspective.

DR. JOHNSON: Rita, this is Dave Johnson. Can I add one thing just to extend the discussion about -- about accountability and how some measures may be helpful, even though they're not really outcomes measures, they are process measures.

The example is one thing that we've been working on is trying to standardize a benchmark for colonoscopy reports, and we worked with CDC and their quality assurance program to come up with a document that says this is a standard that, for example, when a colonoscopy is done, people should do photo

documentation of the cecum and the ileocecal valve.

That's a process measure, but it makes people think when you start to put that in a report that, you know, you documented it. It makes it also recoverable, discoverable and also allows for analyses in quality programs when you go back and you do snapshot analyses of did they do what they said.

And that's a measure that I could see as being a standard, just to make people more accountable and part of the report process would be kind of a quality improvement in and of itself, just holding people accountable, and also allowing for retrospective reviews of these as you start to move into, you know, into a process of evaluating quality report cards.

DR. WINKLER: Thank you.

DR. PACE: Just one comment about real time information on outcome measures, and it's true, if you're talking about risk-

adjusted rates that you can compare to other hospitals or physicians, et cetera.

But there's really nothing that would prevent an organization from monitoring their outcomes real time. So, I mean, you know, an extreme example is mortality.

You can look at each case as it occurs and do an investigation of any issues. If we're talking about function, you could set up some systems where you're monitoring their achievement, and intervene before the patient's discharged if they are not making progress.

So, I think, you know, it is definitely something that providers need to get used to in terms of how they use outcome measures versus process measures, but there are ways to start thinking about how you monitor that real time, and realizing that you're not going to have that risk-adjusted comparison, but if you're comparing yourself to your prior performance, kind of in the

continuous quality improvement vein, you know, most organizations don't have dramatic changes in their own case mix from year to year.

So, risk-adjustment, if you're just comparing your own performance within your institution isn't as big an issue as when you start doing external.

DR. JEWELL: So, I need to ask a clarifying question. Because the call is specifically for outcomes measures, we're not anticipating, are we, that people will have submitted process measures that are linked to outcomes?

DR. WINKLER: No, not as submissions, but one of the things we are going to be doing is going back into that database and pulling out the measures in the topic areas, diabetes.

DR. JEWELL: Right.

DR. WINKLER: You know, we've got probably a dozen, 15 measures.

And so, one of the things we're

going to do is pull those out to look at the process measures, go back and look at the evidence and say, for each of these processes, what's the outcome it's related to, the total list of outcomes, it's just another way of asking what are outcomes for this particular topic area.

So, there are different ways of looking at it. But that was just one approach I was thinking of to help tie all these things together.

DR. JEWELL: No, and that makes sense to me. The conversation, at times, it sounded like what you just described, and at other times it sounded like perhaps process measures that were linked to outcomes would come in as original submissions, and I just wanted to be clear in my head which.

CO-CHAIR DUBOW: The issue of intermediate outcomes. Do you not expect any intermediate outcomes?

DR. WINKLER: No, we do.

CO-CHAIR DUBOW: Well, isn't that what you're talking about?

DR. WINKLER: Yes.

(Off-mic comment.)

DR. WINKLER: Well, not necessarily.

CO-CHAIR DUBOW: Well, it could be a process measure that has a link.

DR. WINKLER: Microphone.

DR. GIBBONS: I think there could be distinctions. I mean, I think for purposes of our work, it would be important to be very specific about the scope.

DR. WINKLER: Yes.

DR. GIBBONS: And I think there are process measures that are not intermediate outcomes.

DR. WINKLER: Right.

DR. GIBBONS: There are process measures that could be intermediate outcomes, and then there's outcome measures.

DR. WINKLER: Yes. At this point

we have kept it as broad as we discussed yesterday, and intermediate outcomes, certainly, as well as functional outcomes, patient-reported outcomes, those all were on the original list and you all kept them on the list.

CO-CHAIR DUBOW: Right.

DR. WINKLER: And said, "Keep them." And so, and even embellished some of them. So, we are casting at that line, but a true process measure, was this thing done, was this process of care, was this test done.

The classic process measures are not something we're asking for or expecting to entertain in this particular project.

CO-CHAIR DUBOW: Okay.

DR. WINKLER: Process measures. Classic process measures.

DR. AMARASINGHAM: Okay. So we are not going to be taking them into account?

CO-CHAIR DUBOW: But we are going to look at intermediate, like blood pressure

control?

DR. WINKLER: Yes.

DR. AMARASINGHAM: Right.

CO-CHAIR DUBOW: Okay.

DR. AMARASINGHAM: Which is a true outcome measure.

CO-CHAIR DUBOW: True outcome measure, right.

DR. WINKLER: Did I hear somebody on the telephone just now?

DR. JOHNSON: Yes. Dave Johnson again.

DR. WINKLER: Oh, okay.

DR. JOHNSON: Would it be, maybe, again, more reasonable to have a little bit of leeway to each of the TAPs to decide really where they think the biggest contributions could be, for example, the discussions we had a little bit yesterday about colonoscopy.

We're going to have a large gap in time until we really have appropriate outcome measures, and we don't even have good process

measures in defining quality and standardization of reporting and things that we, by consensus, the national societies would agree, and we do have consensus documents that would support that.

So, the are not Level 1-A evidence, but if you really want to make a difference in quality, some of the short steps would be process for standardization of reporting, and that's really something that is not at all out there right now.

We talked about things like, you know, withdrawal time and adenoma detection rate, and those, again, are somewhat intermediate outcomes to prevention of colon cancer or reduction of colon cancer mortality, which might take ten to twenty years to show.

So, that's what I'm just seeing potentially more of an issue in ability for GI measures to make a really meaningful difference in overall quality in a shorter time. We might have to have a little bit of

leeway on some of these being process measures.

DR. WINKLER: Well, as we've also discussed, there are areas among the conditions that this project is hoping to address where we realized there just aren't outcome measures yet.

And so, we can't work with something that doesn't exist. However, starting to do some serious thinking about what would be appropriate outcome measures, so that we can encourage measure developers to take them on and so that the next time we try and do this in a year or two, we won't come up empty.

Just as an aside, NQF, this is not the only work NQF is doing. We have any number of ongoing projects that address all sorts of things, and so the opportunity to consider other process measures exists currently and in the future.

So, we're trying to keep the scope

such to address the issues around outcome measures that so many of the stakeholders have, you know, been clamoring for.

And so, that's why this project has the kind of boundaries on it that it does, but realize there are a lot of things happening at NQF. So, it's not an either/or, it's just what are you going to look at as opposed to what is NQF going to look at.

CO-CHAIR DUBOW: So, the opportunity, for example, as Dave is suggesting for the TAP to suggest to some other steering committee or to the NQF staff that process measures are needed in that particular area and may be appropriate, but if it were straight process measure, as we just discussed it, would probably be out of scope for this particular steering committee.

DR. WINKLER: But there are others we could probably --

CO-CHAIR DUBOW: But it could be referred to some other committee.

DR. WINKLER: Correct.

DR. KEALEY: So, will the TAP chairs going to be updated on what other work is going on in their area?

DR. WINKLER: Sure. To the degree you can manage that amount of information, we'll be more than happy to share it with you.

DR. KEALEY: Yes. I mean, it sounds like his impression is that he's working on the latest update of GI measures, where it sounds like you're saying he needs to come up with outcomes and somebody else is working on the latest.

DR. WINKLER: Yes.

CO-CHAIR DUBOW: I think the staff attends all of these meetings and they know how to triage this stuff to go to the appropriate places, so it's not as though every chair has to know everything because the staff provides --

DR. WINKLER: That's our job.

CO-CHAIR DUBOW: Right. That's

what we've been kind of, and the rest of the staff are there for. So, I don't think we're going to lose any opportunity. But, you know, there's a triaging function that will happen. Okay.

DR. WINKLER: All right. Second of the major criterion I think is something that is, again, another thorny one. Scientific acceptability of the measure properties.

And we say that very explicitly because we don't want to go back into the evidence. That's not what we're talking about. So, sometimes, again, a bit confusing in the conversations that we have, but we're looking at the actual measure itself, so it's not, you know, the concept of beta blockers after MI, it's this beta blocker after MI measure and the way it's specified and has been used, and what do we know about it as a measure.

So, the subcriterion, and we'll go

into details and the precision of the specifications, is there ambiguity, are there definitions, could it be interpreted in a variety of ways in different places. That really doesn't help the standardization of comparability of the results. So, precision specifications.

The reliability, validity and discrimination of the measure. Does it work as a measure, is it going to tell us the things we want to know. We are hoping to get information about performance. Can it do it? Is it designed well enough to do that?

Clearly, what we hope is that during measure development there has been testing of these characteristics to find out, is the data that's obtained reliable, does the measure actually measure what you want it to measure, is it valid, and does it, at the end of the day, give you results that allow you to make comparisons.

I mean, that's the whole name of

the game here. And this information is not as easy to come by.

The measurement world has sort of been ramped up in response to a lot of urgency in the market place, and so wanting the measures and getting the measures out, this step has kind of been truncated or at times sort of temporized, if you will.

And so, while I don't want to spend a lot of time on it, we have made provisions for measures that aren't fully tested to the degree we'd like them to, and giving them a time-limited endorsement, for only two years.

But again, you know, that's an uncomfortable place to be, measures that haven't been tested are difficult, and I think we do want to see the degree of what we know about this measure, how does it behave, as we do this.

Question, Mike.

DR. AMARASINGHAM: Just because

I'm new to this process, what degree of validation do we usually require, like, and do we specify the method of validation?

So, for example, can split-sample validation on a single population be sufficient? Do we need to have separate samples? Do we need to sample it on entirely different populations with different socioeconomic characteristics?

DR. PACE: Good questions. For validity, and I think you're talking about the risk model development, the split-sample -- sorry. Are you talking about validity, in general?

DR. AMARASINGHAM: Validity, in general.

DR. PACE: Okay.

DR. AMARASINGHAM: But, obviously --

DR. PACE: I think Reva may have mentioned this at the beginning, and I'll emphasize it now because this is an area where

it really comes into play, and that is that our evaluation criteria are guidance that we don't have, especially in this area, we don't have, like, strict rules like you have to do inter-rater reliability or you have to do criterion validity, and we don't have thresholds, so that we don't have something that says, you know, for reliability, your CAHPS statistic needs to be, you know, .4 or higher in other -- so, what we ask the measure stewards to provide information on what analysis they did and what those results were for, you know, our committee, our TAPs and committees to take a look at.

So, and, you know, we often get that question, and it's hard to just give one answer, because sometimes what testing you do -- or most times, and it depends on what type of measure it is, what the data are that you're using, and so it would be impossible, at least at this stage of our game, to, you know, tell everybody now, if you have this

measure we expect this type of testing.

So, for validity, and the other thing I just want to mention about validity, and both reliability and validity.

So, having precise measure specifications is the foundation for having a reliable measure. And what we mean by reliability is repeatable, reproducible results.

So, if you have those good specifications, that's the first step to moving towards reliability. And the evidence we talked about under importance is that foundation for validity, I think outcome measures have inherent validity, because it's the reason that people seek health care, and the goal of health care.

But, having said that, what we're actually talking about with validity is, can you make valid conclusions about quality of care from that measure.

So, it's a little bit trickier in

terms of validity, and that's one of the reasons that we often don't get information on validity, but quality is kind of an abstract construct.

So, it's not like the validity that, when you take someone's blood pressure you are actually getting their blood pressure. We know, you know, the quality measure of what percentage of patients achieve a certain level of blood pressure is measuring blood pressure are the percentage of patients, but what we're interested in is: does that measure discriminate quality of care.

So, we do make provision, and we recognize that validity is one of the more difficult aspects of testing, and we do say that if face validity is the only validity that is provided, it should be systematically assessed.

So, we would prefer that measure developers, if they are going to rely on face validity, to provide more information about

how that was determined; did they do some kind of voting, some kind of rating among their committee members?

Some of you may be familiar with the RAND method of rating validity of measures. So, having said that, you know, the reality is we often don't get good validity information, and that's why we have this variety of stakeholders together to identify whether there are issues with whether that's a valid measure of quality or not, but this is an area where NQF is, you know, continuing to try to implement and encourage good measurement principles but, you know, you won't always have that information.

CO-CHAIR DUBOW: Dianne.

DR. JEWELL: So, in my mind there's a distinction, at least I thought I heard you say, that there are really two levels of validity. One is the validity of the measure itself which is, I think, what initially you were talking about.

And I would say that in the realm of principles, I would encourage us strongly to ask measure developers to submit evidence of validity of the measure for its intended purpose, because we ran into that problem two years ago with a set of measures that only had reliability data to support.

The second piece of validity, which is, is it valid as a quality metric. Again, I think we're not going to get much evidence, because that's not what they were designed to do originally.

So, not that we shouldn't ask for it, but I think that's where we're not going to find -- where we're really going to struggle for them to submit evidence because they didn't create these things as provider metrics.

DR. PACE: So actually, for reliability and validity, there's kind of two levels, at the data level and then at the aggregate measure, quality measure level.

And so, you know, so sometimes you have for a -- you know, if you're talking about a scale, you might have internal consistency reliability for that scale. In terms of the reliability of the ultimate aggregated measure, maybe not.

There may be some analysis of, you know, what portion of the variation is random versus systematic, et cetera, and that gets very complicated and we have all kinds of experiences with our measure stewards.

So, it will depend on, you know, what type of measure, what the data are but I think, you know, we need to think of these as building blocks, and so we may not be at the level of the most sophisticated testing, but we'll encourage everyone to be looking at, you know, first is it on a sound foundation.

Because, if those first two foundations aren't there, the chances are you're not going to get to a measure that would test out properly.

But, you know, it's something that, you know, you all will have to work through because it's challenging.

DR. AMARASINGHAM: Just a quick follow-up. I mean, the conundrum that I see, and I'll propose a potential way of looking at it is that, you know, on one hand we don't have enough measures.

And so if we keep waiting for a certain level of evidence for standards, we're not going to ever have any measures. It's going to take a long time.

On the other hand, I'm concerned, as someone on the ground taking care of these patients, that among my colleagues, there's always this concern that these half-vetted measures come out that ultimately, after five years you find out aren't very important, or aren't validated appropriately.

And so I wonder whether there's some middle ground of a level of confidence in the measure that, you know, if a measure had

not only face validity, but criterion validity, discriminant validity, convergent validity, concurrent validity and has been tested in multiple different populations on both national and local data sets, that's an incredible measure.

DR. WINKLER: Yes. Have you seen one? Do you know of any?

DR. AMARASINGHAM: Well, I mean, for example, you know, I thought that the hospital readmission is on the way to becoming a very good measure. I think it has some serious flaws in certain areas, but among measures, that measure was very well done, done in a Connecticut sample, done in a national sample.

They did c-statistics. They did, you know, large technical papers on it. A measure that just has face validity would have extreme concern to me, especially if the NQF measures are tied to accountability and pay for performance, and the average person can't

distinguish which measure should be used for which, and if it comes out of NQF sanction and potentially is in widespread practice, what you have the potential of doing is saying the people that are deciding this really don't know what they're doing and I can't distinguish between the most important measures as a person on the ground.

And so, I think the NQF, if it hasn't been considered already, needs to think about something like the US Task Force for Preventive Services or others that kind of have grades of evidence and levels, because, you know, I think there are certain measures that are phenomenal.

At the same time, you don't want all measures to go through that process, because it's going to take ten years, how you get practice with the measure and so forth.

And because I think these very careful levels of validity need to be demonstrated for each measure, and if we don't

have it, but we believe a measure is on its way, if we intuitively believe that a measure could be very useful, you know, then we should present it as a certain level of confidence and, you know, I mean, because I think there are important measures, but I just don't think we're going to have that level of evidence.

DR. PACE: You know, it's an interesting point and it has been brought up. You know, right now NQF's process is endorse or not endorse, or time-limited endorsement for measures that are not tested.

We do ask, you know, the reviewers to rate each of the criteria or subcriteria, but pretty much on a scale of kind of completely met, partially met, minimally, or not at all, in helping you come to a conclusion about recommending or not recommending.

But it's certainly an area that we need to be thinking about in terms of whether we want to institute some kind of grade to the

endorsement, is what you're saying, you know, and it's something that we would need to discuss with our CSAC and ultimately with the Board to institute something like that.

But, you know, certainly, throughout this process, you know, to certainly think about that and, you know, that can help us, you know, sort out how something like that would be operationalized.

At this point what the Steering Committee has the option of doing, and Helen can chime in here, is you know, as I said, it's either recommend or not recommend, but the report can identify any specific guidance that the Steering Committee wants to at least make known in terms of your decisions.

The reality is that we don't have control over, you know, how measures are implemented, but I think that's certainly something we should continue to think about.

Helen.

DR. BURSTIN: I apologize for

being late. I was on a safe practices steering committee and we were discussing the grading of evidence. That's just the story of my life.

So, this is an important issue. It's not going to go away. Clearly, we need to kind of reconcile this. Whether we would actually grade the measures, per se, or actually have a more formal assessment of the grading of the evidence underlying the measures, that's more transparent and easily reconciled to something I think we need to do a better job of.

The US Preventive Services Task Force which I oversaw at AHRQ for five years doesn't always fit many of these kinds of measures that grading evidences. It is really two grades, which I think people often forget as well, is actually the grade which is the overall recommendation of, you know, an A recommendation, the benefits significantly exceed the risks, all the way to a D, don't do

it, risks exceed benefits.

But there's actually a second grading system of the quality of the evidence. I think that's kind of what we keep hearing as a recurring theme, is it's less about the overall recommendation, A, B, C, D, E -- A, B, C, D, I -- no E, but instead the recommendations are on the grading of the evidence as being a more crucial input we need to be more thoughtful about how we grade. It keeps coming up.

DR. PACE: But I think what he was talking about was specifically how confident we were in the measure.

DR. AMARASINGHAM: The only point I would say is that I've been impressed with the way CMS has done some of its measures in that there's sort of a pilot period of two years where everybody's getting use to the measure and actually quite a bit is learned about the measure, but if that sort of semi-sanction hadn't come from CMS, no one would

have tested it.

So the question is: Would there be a group of measures that NQF says doesn't meet our set of gold standard level of evidence, but that we would encourage regions to experiment with accepting, and I bet that might take hold, because there are places that would like to experiment.

DR. BURSTIN: And the only thing I'll add is, I mean, that was somewhat of the thinking of the idea behind having a designation for time-limited endorsement.

And I think you could certainly say the CMS measures that have come to us are very well tested. They have been extensively tested. Well, what they don't necessarily have, which as CMS enables, is a field test.

DR. AMARASINGHAM: Right.

DR. BURSTIN: Where hospitals and others have a chance to see the results and reflect on them, and that's a question of, as we sort of get a better sense of the

performance of some of those measures that got time-limited endorsement, we may reevaluate what field-testing really means in terms of how this all fits together.

CO-CHAIR DUBOW: Iver, did you want to say something?

DR. JUSTER: Yes. I was actually going to ask whether this time-limited endorsement was stamped clearly somewhere so that whether it's report cards, P4P, whatever, that people wouldn't think that these measures should -- P4P or public reporting, at least not on the same list as the ones that were endorsed with higher validity.

My second question was whether in the portfolio you have examples that could easily be shared with this group of what would be considered -- they did a really good job of validating this measure.

They did an okay enough job considering the kind of measure this is, and the other one, well, the other ones, I guess

would be the two-year ones, be pretty clear which ones those were, so that we can stand on their shoulders, basically.

DR. WINKLER: Yes. Excellent suggestion. There's absolutely no reason we can't pull those out of the database. Karen's done so much work on outcome she could probably come off the top of her head and come up with a couple.

And it would be -- yes -- no, but the good, the bad and ugly, I mean, we can do it all. So, we'd be more than happy to share that with as examples. Just try not to get them confused with the work that's being asked of you to act on. They are strictly a reference sort of thing. We'll do something to make them look not actionable.

DR. JUSTER: Okay. Yes, the question of validity, you know, if you're thinking, I have a new test, I have a new imaging test that's a lot safer than a pulmonary angiogram, but I already know that

I already have an idea of what the gold standard is, so I assume that's a hundred percent accurate, and now I have this new test and I'm going to compare it to the gold standard.

Well, we don't really have that a lot here, so we have other ways to get at validity, I think.

DR. WINKLER: Your question about is are they stamped time-limited. NQF stamps them on the things we have control over. Once they get out into the field where we don't have control over things, it's kind of a jump ball.

Some people are fairly good at it and some are not. So, you know, that's sort of our influence only extends so far.

DR. AMARASINGHAM: That's where the recommendation would be so valuable.

DR. WINKLER: Yes.

DR. AMARASINGHAM: Because if I was a person in the field using this and I

said NQF says this is so good it could be used to determine your reimbursement, that to me is a level of evidence that's a little bit different than we have some early experience with this measure, and you could use it for quality improvement at your institution.

I mean, I think it would be valuable, particularly for us as we're making decisions to be able to have a kind of framework like that.

CO-CHAIR DUBOW: But it should be clear, NQF does not get deeply involved in the implementation of measures. So, to Iver's point about whether, you know, we say this is okay for pay for performance, what happens with these measures post-endorsement is not within the purview of NQF.

So, I understand your point. If we had that kind of designation, people would understand that it had exemplary properties and would be suitable for something of that sort.

DR. AMARASINGHAM: Because I'll say, as a committee member, I just decided at the last point, as a committee member that, you know, if it was my level sort of a methodologist, my level of standards, a lot of measures I would say no for NQF.

That's why I'm wondering whether there could be something beyond an all-or-nothing standard, and I think different people in the room would have different standards. But I would say, as a strict methodologist, I would not approve most of the measures.

DR. WINKLER: Yes. You've kind of hit the crux of NQF, if you will, in those last two statements, because one, it's a multistakeholder organization, and so the levels vary everybody, and that's what this is, is a negotiation.

Your suggestions about levels and how measures are used is not a new one. We've heard this many times in many different venues. Again, that would be very hard to do.

And it's not that we haven't thought about it, we just haven't figured out to do it particularly, that suits our multistakeholder kind of approach to things.

So, at this point it's not that we haven't considered it. We certainly have, but at this point it just isn't working, it doesn't seem to suit the current construct of the way the NQF is organized.

But we keep thinking about it.

CO-CHAIR DUBOW: Pauline.

DR. McNULTY: Yes. Just when it comes to listening to this debate, but when it comes to the patient-reported outcome measures, I think I've mentioned this already, there is a draft guidance from the FDA out there which is considered a best practices on the development of patient-reported outcomes measures and testing them in terms of reliability and validity.

So, I think you really should look at that, and I'll send you the link for that.

But, you know, one thing that, you know, just keeps popping into my mind as I'm listening to the discussion around validity, I was at a meeting a couple of weeks ago and I heard somebody who used to be at the FDA talking about validity and reliability, and the comment that he made really just stuck with me which is that, you know, you can have great reliability, interrater reliability on some kind of measure, I mean, it's mostly scales that I would be dealing with, and you could say it's a reliable measure. However, maybe the validity isn't there because the thing that people have been asked to rate is that the moon is made of green cheese.

So, you know, that's not valid, yet you can get great interrater reliability if you got everybody in the same room to agree on it. But you still don't have a valid measure.

So, it's kind of a perfect thing to keep in mind about this debate that we've

just had around validity.

DR. PACE: Right. And that's an excellent point. Reliability is necessary, but not sufficient to prove validity, and so you can have something that's absolutely reliable on giving you the wrong information. Right.

DR. HOPKINS: So this conversation reminds me of so many we had at the CSAC and elsewhere and every steering committee. It's good.

But, I keep thinking that our focus here is a little bit different from what it's been in these other NQF committees, because we're talking about outcomes, so I don't think our job is to figure out if this process is linked to some outcome. If anything, it's the other way around.

We're supposed to be focusing on the outcomes, and as I look at the list here, it seems to me like we need to think a little bit more about how to apply that list which

was really constructed more for the process measures to straight outcomes.

How do I judge validity, 30-day mortality following X? I mean, I understand all the complexities of how you adjust for risk and all that. That's a separate item.

DR. AMARASINGHAM: Well, I think for that one, I think the big question would be the risk adjustment. But the other -- right.

DR. PACE: Right. And so, with an outcome measure, risk adjustment affects validity, and so we may not need a separate test of validity, but we do need the risk adjustment or the issues that threaten the validity of the measure.

Exclusions can threaten the validity of a measure of quality. So, you're right that, you know, we may need to look at these different, but the concept of validity is what we expect for testing may be different, depending on the type of measure.

DR. JEWELL: Well, and also the issue of validity for the self-report measures is absolutely relevant to outcomes, you know, when you're looking at what pay -- I mean, that gets to the point you just made.

If I'm looking at disability or looking at function and I'm doing that by way of a patient self-report questionnaire, the issues of validity are absolutely something we would wrestle with.

DR. PACE: Right.

DR. BURSTIN: Just one other comment is that I think that in general while we have had time-limited measures without testing that are mainly process measures, outcomes tend to take on a higher level, and it's almost like it's sort of inconceivable that a nontested measure would likely come forward to this committee for an outcome.

So, I think that probably won't be an issue.

CO-CHAIR DUBOW: But we won't have

time-limited?

DR. BURSTIN: I suspect we will not. I think it's very difficult to really ensure validity of an outcome without having any testing of any kind.

DR. JUSTER: Yes, and I would agree with that, that certainly process measures, I could imagine somebody saying, well, did you have an ACE inhibitor in the last year -- well, I guess that's not a terrible measure, but it's a process measure, but it doesn't discriminate very well against people who will have a good outcome because they took their ACE inhibitor every day, than those who filled it once.

But even for outcome measures, one might consider, for example. And I think it's on here, this discrimination thing, so suppose you have the SF-1 we were talking about yesterday. How are you feeling, and it's a five-point scale, that might not discriminate quality of care very well in the sense of you

have to move so much to move to the next point on a five-point scale of a one-question item, that no matter how valid the measure is, it doesn't discriminate quality of care very well.

There might not be very many interventions that could move a population by a whole scale point. Is that what you're getting here with discriminating quality of care? Nice measure, great outcome, but it just doesn't discriminate quality of care very well.

DR. PACE: Well, again, we're talking about measures at the aggregate level of the provider, so what we would want to see is provider-level scores, whether it's a hospital, a physician, a home health agency, where there's some discrimination of quality.

So, you know, that's the whole point of these measures, is to identify where improvement is needed for quality improvement, or to identify providers that you'd want to

seek care from.

And so, if you have a measure that ends up where there's no distinction among the scores, we're doing a lot of measuring with no next step to happen from it, so that's what we're trying to get at.

DR. JUSTER: Sort of like the old satisfaction surveys where almost everybody was either satisfied or very satisfied, it doesn't discriminate much. In this case I'm thinking that the scale might, no matter how valid it is, having a provider move their entire practice by one scale point on a five-point scale would be herculean.

CO-CHAIR DUBOW: I just want to point out to those people in the audience who want to say something, we'll have a public comment period about five minutes before our break at 10:30.

Okay. Reva, we haven't gotten to exclusions yet.

DR. WINKLER: Thank you.

Exclusions are in red because this is a topic that has been pulled out specifically because we've had to struggle with it over the years with the issues around exclusions.

A couple of things. Issues increase the complexity of measurement burden. You have to collect more data, and the more exclusions, the more data.

Often, the exclusionary things are hard to identify so that they're not necessarily in maybe more readily available data streams.

They often create a barrier to measure harmonization, and of this beta blocker measure and that beta blocker measure, one excludes this and another one excludes three things or not the same or three different things, so that the measures can't work well together as a group.

So that we really want to see evidence that the exclusions that are part of the measure are important parts of the

measure. They actually contribute something, that it would be distorted without those specifications.

Also, if patient preference is a consideration, the numerator or the denominator exclusions, it should be specified so that the effect of the patient preference on the measure is transparent.

And the classic one we had to deal with, and Karen can take you through this one, is flu vaccination rates and patient refusal of flu vaccination.

How do you accommodate for flu vaccination? Do you get rid of them in the denominator or do you include them in the numerator as a separate category. That way you know what happened to everybody. So, you know, there are a variety of ways of doing it.

And, Karen, did you want to talk about that particular --

DR. PACE: Yes, I think, yes. That was a perfect example, because there was

some concern about just removing patients that refused from the denominator because then that information just goes away, and you could have providers with identical scores, but one had 50 percent of their patients refusing, and another one, you know, with the same score with all of their patients included.

So, the approach we took in that project, which was a big harmonization project, was to make that a numerator category so it would at least be transparent, so that we would have the actual rate of immunization, but that that could be reported, as well.

I think probably these come up, the exclusions tend to be more of an issue with some of the process measures, but having said that, if you remember back, our conversation yesterday, that one way of having an outcome measure that maybe doesn't have a sophisticated risk adjustment or even stratification model is to have a narrow homogeneous patient population.

So, then exclusions will play a vital part of that measure and, you know, need to be carefully looked at by view in terms of, you know, what we're accomplishing with that kind of measure.

So, it is something that we want to pay attention to.

DR. WINKLER: Right. The devil's in the details.

DR. HOPKINS: I just have one piece on that inclusion for patient preference. There's also a feeling among some of us, and I think some evidence to support it, that some clinicians are actually more effective in getting patients to do what is good for them.

And we didn't want to lose that in the measurement of quality.

DR. WINKLER: The area of scientific acceptability, you know, is sometimes fairly thorny, fairly scientific and fairly beyond what I understand. I'm very

happy Karen's a good friend and colleague.

And so, I just want to remind everybody around the table that if you got lost in that conversation and it got a little too down in the details for you, remember that the evaluation of these measures is a team effort, and there is a very deliberate reason we have different areas of expertise around the table.

And I think you'll find that, as we go through the measure evaluation criteria, your particular expertise you bring to the table will feed into different elements of it.

And so for those of you who are methodologists, really enjoy the reliability and validity discussion. But for those of you who are more in the audience kind of realm and we're going to talk about usability.

Is it useful to you, does it give you something you want to know, let's move on. So, it is deliberate, and that's why it's going to take all of us to come to a

reasonable conclusion on how to recommend the measure go forward.

Usability. What we'd like to see is evidence that the measurement results are meaningful and understandable to the intended audience. It's like, you know, you can create all sorts of information, but does it mean anything to anybody, is it actionable, is it useful, does it respond to the needs of the audiences for information.

And this is where particularly the various stakeholders have a real significant role to play is, is it going to be useful for the consumers in terms of information about health care.

Is it useful to purchasers to understand the value of the health care they're purchasing? Is it useful to health care organizations and institutions to understand and be able to improve the quality of care they provide? Is it useful to professionals to understand the performance

that they're providing?

So, if data is just data unless it's useful. And so, we really want to avoid the, you know, just data, and is it useful for a wide variety of audiences.

This is why one of the criteria is that the measure is useful by the evaluation criteria, by the intended purpose of the measure developer when they submit it, is that it is not just for internal quality improvement, not just, you know, for fixing things in your own house, but that it is suitable for public reporting and useful for a wide variety of audiences.

In addition to usability in trying to enhance that and to make it as easy as possible for measures to be used by purchasers, reporting systems, providers systems, whomever out there, is that the measures are harmonized.

So that we've got four or five different diabetes measures from different

places, but diabetes is defined differently in all the denominators. Oh, you know, that does not help anyone.

If we can get the common definition of diabetes, then all those measures can work together to provide a much more robust picture of the performance of what's going on, but even though they came from a variety of places.

So, harmonization is becoming a real critical issue for usability because if it's not harmonized with the measures you're already doing, you're probably not going to adopt it.

However, if it is, and you can easily fit it into your portfolio because, hey, we're already collecting data on all those diabetics, we'll just, you know, pick one more numerator data point. Fine, we can do it.

So, usefulness, if things are just so out of step with everything else that's

going on, it's just going to be that much harder, the barriers are greater to get them implemented.

So, usability again. This is where your stakeholder perspectives becomes very, very important, because useful to you in the world you come from and bring that representation to that committee.

Because I'll tell you, when we go to comment, this is an area that we hear about lots. I don't know what in the world I'm ever going to do with this measure, you know, from a variety of the stakeholders.

So, this is really an important area for you, and it sometimes doesn't get the attention it needs, so I really like to emphasize it, and don't be shy about bringing your concerns forward about utility of these measures.

Questions on that particular criterion?

DR. DEUTSCH: Just, can you give

us an example of a measure that was not for public reporting, but just quality improvement? You obviously have a reason why you put that there, and I'm just struggling with an idea that might be appropriate for us.

DR. WINKLER: You want to use.

DR. PACE: I was just going to say that in the earlier years of NQF, we had some measures come through that were developed primarily for quality improvement, and were endorsed for quality improvement, and since then, you know, through policy, and now more explicitly in our evaluation criteria, we say that the measure should be intended for both purposes, both public reporting and quality improvement.

And we probably can drop that kind of highlight, but this was, you know, a very explicit, wanted to call it out as something that we were emphasizing.

DR. BURSTIN: Just to add to that, I think that part of our thinking is also that

there are so many measures out there, even beyond the hundreds we've already endorsed that are really very appropriate for internal QI, but you would not want to publicly report those measures necessarily if they don't, in fact, achieve the same level of quality of the measure itself, that we would want for a public-reported measure.

So, the measure that's useful for QI is great, but we also want to make sure it's also appropriate for public reporting.

CO-CHAIR DUBOW: This has been a subject of ongoing discussion and debate. Ongoing meaning over a lot of years. So, the fact that it's here just was trying to put it to rest to clarify purpose, it has to be dual.

MS. GERBIG: Just an actual example of something like that might be, to prevent central line bloodstream infections, you could measure the five step process that prevents it, but many of us measure the number of days or the number of years without a CLPC

in our organization.

Now, that's not a publicly-recognized number. It would be a doozy of a number to try to ever report publicly, but that's sort of an outcome measure.

And just from a user's point of view, that's why I'm so interested in the work of this group because we have all of these process measures, but they don't answer the question, so what.

And the outcome measure answers the question, so what. And so, in a perfect world you'd have process measures, but always an outcome measure that is the overarching measure and allow perhaps providers some wiggle room to implement the process measures with keeping your focus always on the outcome.

DR. WINKLER: Linda David just whispered in my ear. She gets it.

MS. GERBIG: I live it.

DR. WINKLER: So, anyway, any other questions or comments on usability? It

is an important criteria, but sometimes we always get lost in the discussion around the science and the validity and, you know, that goes on for hours, and the usability often gets short shrift.

So, don't let it, because it will ultimately -- it's the endgame, if you will. If it isn't usable it can be as valid as it wants to be.

Feasibility. Again, often a topic that engenders a great deal of discussion, and feasibility, again, there's a wide variety of what's feasible out there in the world. I'll just give you an example of something that I've lived with, Alexis and I have lived with for the last year, and that's our project in clinically enriched administrative data.

In looking at the measures that came across in that project, we had 206 measures submitted, so you know, sometimes a few measures is a relief, flooding is not always appropriate.

We looked at the measures and they have a wide variety of characteristics, and you know, we basically were able to categorize the measures into a couple of different levels.

And one is a level that has just a single, like, traditional claims data stream, basic, pretty much anybody can do it. All right.

So, on the level of feasibility, real high. The problem was that measures are not real robust and they tend to have the criticisms that many people have with the straight claims measures.

Second level were measures that pulled together two different claim streams. It's like medical visit claims, pharmacy claims, lab claims, whatever, but requires methodological complexities, you needed to have some people who could do this stuff and combine the different data elements together.

But again, fairly straightforward.

A lot of people can do it. Not everybody, but still a lot of people can do it. So, while this one is high feasibility as level one, it's still a reasonable level of feasibility to do.

We get to the third level which is really what a lot of people think clinically enriched means, and that is, to one or more claims streams of data you add electronic clinical data. The most classic example is laboratory values. Okay. Not just did it -- was it done, but what was it.

And we have a small number of those sorts of measures in the portfolio. Those measures, very robust, measures everybody loves to see, but the feasibility suddenly drops because you just don't have lots and lots of organizations that are able to do it.

But yet, hopefully in the future, more and more organizations will develop the capability to manage data in that fashion such

that feasibility can be improved.

And the Steering Committee took the stance, of very deliberately choosing to have some measures in level three and some in level two, hoping to sort of point the way, encourage more measurement complexity and more measurement rigor.

So, feasibility can become a really cornerstone of a project such as it was in that last one. And so, your expertise coming from whatever realm and world you live in, understanding is this a measure I'll be able to take home and do, can this be done at my house, without undue burden. How much burden will it encounter?

And so, that becomes a factor as you mix in and understand all these criteria. There are no thresholds, there are no absolutes, but it's a factor because it really makes a difference how far this measure can be used and will be used going forward.

So, feasibility becomes a really

important part of the assessment. Corollary to that is realizing the work that we and a lot of people are doing to transition us to an electronic world, and the feasibility has, you know, the ability to maybe translate some of these data elements and make this ready for moving into the EHR world, again, is another aspect of feasibility that we want to see.

You know, the classic old chart-based measures, you've got to pull paper records and read pieces of paper and abstract it onto a form. I mean, those are probably the least feasible and most expensive to measure, kind of measures known.

There's a whole gradation of feasibility. So, an assessment on the feasibility of the measure, who can do it, how much will it cost them to do it, and I don't mean just in terms of dollars, manpower, time, resources that might be otherwise used somewhere else, to do that.

So, certainly with an emphasis on

electronic collection, if they're still relying on some sort of hand collection, you know, what's their plan for losing that, because it's just not going to make it very feasible going forward.

So, Karen, did you want to say anything about feasibility?

Helen? Like I say, it's the world I've lived in for the last year.

Joyce, David, feasibility, anything to say?

DR. HOPKINS: You said it well.

DR. WINKLER: Okay. Alrighty. So, what we want to do is give you an example and we sent one to you in your bundled set of things. I don't know. Which one did we send them?

DR. PACE: Yes, I just want to mention, you know, we sent you this early example just to give you some idea of what you'll be seeing, but please keep in mind that this particular example, the measure steward

had informed us, and we knew this in advance, that they were still completing some of their analysis so this is going to be updated, so this is not the final information.

But we just wanted you to see an example of some of the information that you'll be getting. And we'll also be working with all of the measure stewards if we want information moved to a different section.

So, this is just a brief look at the types of information you may be seeing, but keep in mind this is not the final, this measure is going to --

CO-CHAIR DUBOW: Just before we go to the example, it is 10:30, and I did mention that we would allow for public comment, because I assume that you still want to make a comment?

MR. HARDER: Yes, I do.

CO-CHAIR DUBOW: Thank you.

And, Operator, if there is anybody on the phone who wants to make a public

comment, could you open the phones and ask, please, we have one comment here in the audience.

OPERATOR: All lines are now open.

MR. HARDER: Great. Can you hear me? I wanted to go back to the risk adjustment models and just highlight that. Please be aware that sometimes there's going to be two risk adjustment models for the readmission measures and think about this.

It's going to be based upon planned procedures, which are done for the sickest patients, and I wanted to emphasize that publicly reporting both of those is going to be a concern in our case because we think people don't understand the rationale behind the plan procedures.

One thought is that we're doing it for the money, you know, because you get double amount of the money, but also you've got to realize that these are the sickest patients and that this is in the best interest

of the patient because they can't handle the contrast or they can't handle some of the stresses of the procedure.

So, I just wanted to bring that up in that discussion about scientific acceptability and validity. Thank you.

CO-CHAIR DUBOW: Thank you.

Operator, is there anybody on the line who would like to make a public comment?

OPERATOR: All lines are open.

CO-CHAIR DUBOW: Okay. Thank you.

DR. PACE: Are we going to take a break?

CO-CHAIR DUBOW: Well, I was just going to ask. Does everybody want to take a break? Yes. Okay. We're going to break for 15 minutes, and then we'll come back and have the example presented.

(Whereupon, the above-entitled matter went off the record at 10:28 a.m. and resumed at 10:51 a.m.)

DR. DEUTSCH: I just wanted to

kind of wrap up on the previous conversation, if we're getting started. Just a couple of questions.

I guess one thing is, Dianne and I were just kind of talking break about the issue of, you want to harmonize, we want to have measures that work across diagnoses, but we also want validity.

So, something like readmissions, certainly what would be preventable as readmission might be different by different diagnoses. So, would we expect to see one measure that had just risk adjustment or different exclusion criteria or is that --

DR. PACE: Well, I'll just make a comment, and then others can chime in. We have quite a few readmission measures. The ones we currently have endorsed are primarily all-cause readmission.

One of the issues about the preventable readmissions is getting into agreement about what's preventable, and the

methods that have been used to date to try to classify those.

And so, you know, if we do get measures of preventable readmissions it would have to be specific for a particular condition in order to go through that process of identifying what would or would not be preventable.

DR. DEUTSCH: Okay.

DR. HOPKINS: Okay. So, here's another example that I've been thinking of. So one of the focus areas is cancer. Almost every cancer researcher uses as a fundamental measure of outcome disease-free survival, you know, for X years.

Are we going to have to go through a process of approving 20 of those or however many subcategories there are within cancer or is there some way we can arrive at a measure which is disease-free survival after five years for cancer patients?

And, how will that play out,

because I'm trying to track the flow. You know, the call for measures. Who owns that disease-free survival measure? That's the first problem. If we don't come out of here endorsing a measure like that for cancer care, it really blows my mind, but I don't know how you're going to get it.

DR. WINKLER: David, I think that has been one of the measures that has been used sort of on both within the research realm, the clinical realm and the public health realm, and all of the, you know, the cancer world and registries as sort of a data point. But I'm not sure it's ever been really thought of and portrayed as a quality measure. It's information. It's an outcome, absolutely, but from a quality measure perspective.

So, I think it's an interesting challenge because certainly I'm not hearing it and I have yet to see anything submitted like that, and I'm not really sure I expect to.

What do you think, Helen?

DR. BURSTIN: I actually had a conversation with Jane Weeks, who's at Dana-Farber, probably one of the guru's in outcomes research and cancer specifically saying that.

So, what are the outcomes out there, and her response back was that that's actually a really difficult measure to track because there are so many complicated issues around this issue of the different kinds of diagnoses and things like that.

I have to share it with the group. It was a very thoughtful response. And, again, this is the kind of thing where, hopefully, I mean, we have a really strong chair in Lee Newcomer who really understands this issue well.

So, you should talk to Lee in advance, or see if you have specific concerns. But those measures currently are not used for public reporting.

CO-CHAIR DUBOW: And don't forget,

there has to be a measure that comes to us to evaluate.

DR. HOPKINS: That's what I'm worried about. That's exactly what I'm worried about.

CO-CHAIR DUBOW: So, you know, that measure that you like so much may not exist.

DR. HOPKINS: Well, it exists throughout the research community.

CO-CHAIR DUBOW: Well, is it a numerator, denominator kind of measure?

DR. WINKLER: Yes, it is.

DR. HOPKINS: I think so. I mean, it's very straightforward.

CO-CHAIR DUBOW: Well, then bring it on.

DR. HOPKINS: I have to find somebody who owns it, that's the problem.

CO-CHAIR DUBOW: Well, there you go. After you finish your --

DR. HOPKINS: But I'm thinking of

the patient, right. What does the patient want to know. That's fundamental.

DR. PACE: But, I mean, my concept would be that you would need something diagnosis-specific, because those disease-free survival rates are very different, depending on the type of cancer, and if you're going to try to make some assessment of whether your center, you know, is treating appropriately -- anyway, there's a lot of questions that you raise.

DR. YAWN: One of the ways to think about that is quality of life, but you have to be very careful of what you put in the denominator and how do you assess someone's quality of life if they are not alive?

And so, there's all kinds of fascinating ways to look at that, and some of the cancer survivor papers, they do try to address some of that, and assign a zero quality of life if you're not alive.

CO-CHAIR DUBOW: Okay. So, let's

get back to - yes, it will certainly be continued.

DR. DEUTSCH: Sorry, one other quick comment. I'm not sure where this fits in in terms of the criteria, but just kind of potential unintended consequences might fit under importance.

I kind of heard, you know, benefit and just as an example, falls. I mean, you already have items related to falls, and my concern would be if you're encouraging and falls are not good, but you might encourage that the patient would stay in the bed and become debilitated, or not get all the care that they needed and not get up and around.

If the staff are so worried about falls, discharge to community is something we measure in rehab, but is the person really ready to go home. And so we don't want to encourage people to discharge somebody home.

DR. BURSTIN: Just one follow-up point to that, and I see lots of heads

nodding. I mean, one of the things that's under, I believe it's feasibility, is unintended consequences.

So, it's a really important consideration, this issue of, you know, the catheters versus falls versus taking out the catheters. I mean, there are just trade-offs in so many of these things, and the last thing we want to do is have unintended consequences because of measurement. So, we really look to you for that thoughtful commentary. Is your head nodding? As a hospitalist this is reality for you.

CO-CHAIR DUBOW: All right. We're going to now go to just sort of take a high walk through the evaluation sheet for the AMI mortality, just to give you a sense. It's not completely filled in. We did receive a copy of this, and Reva and Karen are going to --

DR. WINKLER: One thing just to wrap up our earlier discussion, in your materials that was sent to you, here is a copy

in hard copy of the measure evaluation criteria laid out in all of its glorious detail. And so, for things I may have not gone into appropriate detail on or whatever, here it is all laid out with the subcriteria, each one, if you notice, importance to measure and report. You know, there's criteria, whatever it is, I can't even read. Yes, where's my eye outcome test. It's 1-A, 1-B, 1-C and so through all the various criterion.

So, there it is in detail for you as a reference. And this is an important document for you to keep at hand as we go through the measure evaluation process. So, I just wanted to point that out to you. You don't have to rely on remembering what I said, or even understanding what I said, if I didn't say it well.

Now, we're bringing up a version that you have this measure submission form as an example in your packet, but it is a PDF form, and what we've got is an example of what

you're actually going to receive and this is, again, Karen's work.

What we've done is place the information that got submitted. We've reformatted into a measure evaluation tool form with the information. So, the blue information that's in this form is what the measure submitter, the measure developer gave us. They entered it into the appropriate question. It gets dumped in here.

One of the things we're noticing is people are putting their answers to the wrong questions in the wrong spot. That's always fun, and incomplete, not answering things. So, this is an example of things that aren't completely filled out, but it can give you a sense of that.

And Karen is going to walk you though it because there are some characteristics of it that she put in this tool that we want you to be aware of so that you can use it optimally.

DR. PACE: Right. And as I mentioned, we knew in advance that this particular measure wasn't fully complete. The measure steward had notified us of that, and they're ready to submit the rest of the information. So, we'll be getting you the final information when that's available.

So, as Reva said, we're importing the information that is submitted online into this form and this form has embedded in it the evaluation criteria.

So, I'll point out a couple of things. There are some areas for NQF staff use so that we'll make sure that certain, you know, the numbers are in there, the NQF staff will be checking that the conditions are met before this measure even gets to the TAPs or steering committee.

And so there's some color coding in here in terms of the gray-shaded areas or NQF staff. The yellow-shaded areas will be for the TAP work group review and the pink-

shaded area is for steering committee. And the reason for that is, as Reva mentioned earlier today, what we're going to be asking the TAPs to do is to evaluate each of the subcriteria and provide their advice back to the steering committee in terms of how well they think the measure met those subcriteria.

They will not be evaluating overall that big criterion like importance or scientific acceptability. Their assessments will be provided to the steering committee who will make those bigger evaluation comments. But I'll get to that in a minute. So, here you'll see that if the conditions haven't been met, the staff will make some notes back to the steward and send it back.

If the staff have any particular notes to the reviewers, they'll put them in here. If there are any particular questions or issues they want you to be particularly aware of. There's a place for the staff reviewer name, the TAP, and the steering

committee names.

Our intent is to try to build this as we go along so that we eventually will have all the information, kind of the summary information from the TAP and from the steering committee, not each individual reviewer's information.

Okay. So, in each section, there is a link back to the evaluation criteria, so if you're looking at this on your computer and you have internet access, if you want to go back to the criteria, this will be a link back to the web page for that. But also embedded within here are using the comment function of Word, and if you move your cursor over -- why is it not staying up? Let me go to another one and see if it will -- it was doing it before.

Yes. But anyway, the actual criteria language are embedded in there, and -- I don't know why, it was doing it before, so we -- yes, we've had all kinds of gremlins

with computer technology.

The other thing is, you know, with your own system, if you prefer to have those comments in balloons along the side, if that's easier for you, that's a possibility as well.

But let me see if I can get back - -

DR. WINKLER: What we're trying to do is make everything in one place so you don't have to be flipping pages and going back looking for documents. We're trying to embed all the information so you don't get lost like I tend to do.

But, in terms of the rating scale that we'll be using, again, the TAPs will be using it for the subcriteria and the steering committee for the overall criteria is kind of a four point scale. C means completely met, unquestionably demonstrated to meet the criterion. Partially, P, partially, M minimally, and N, not at all, or incorrectly addressed. So sometimes you'll see stuff

filled in, but it may not really be responsive to the criteria or the question.

And then there are some areas that are NA not applicable, for example, justifying exclusions if there are none. Then, that would be, you know, a valid response, not applicable for that particular measure. So, where not applicable is truly not applicable, that would be an option. Okay.

DR. McNULTY: Karen, can I just ask a question. The piece under measure descriptive information is the National Priority Partners Priority Area. Is this filled out by the measure developer?

DR. PACE: The priority area is filled out by the measure developer, because that's just a categorization and those are just those six broad areas. But if you go down in importance there's a section where the staff will fill in the specific goals.

DR. McNULTY: Okay.

DR. PACE: So, right here, for NQF

staff use, they will be filling in -- they'll be checking to see if the measure addresses one of the specific goals and if so, they'll be putting it in here so that you'll have that information when you get the form.

DR. McNULTY: Okay. And then in that same section earlier where the developer fills it in, consumer care, need, getting better, is that terminology that you use or the developer just --

DR. PACE: That terminology is from consumer language and it's terminology that --

CO-CHAIR DUBOW: It was part of the framework that was used to present information to consumers, and so they -- I can't remember, was staying healthy --

DR. PACE: Yes.

CO-CHAIR DUBOW: Getting healthy.

DR. PACE: Staying better.

CO-CHAIR DUBOW: You know, living with illness.

DR. PACE: Right.

CO-CHAIR DUBOW: There were four, weren't there four?

DR. BURSTIN: End of life.

CO-CHAIR DUBOW: End of life.

DR. PACE: Right. And I should also mention that most of the things in here are things that NQF asked in our last round of measure submission, but since that time, Reva led a project for data fields collaboration with a group of measure developers. And so we've come to try to reach some agreement on the types of information, and that was one that one of the other groups was, I think, AHRQ - was using to categorize the measures in the National Quality Measures Clearing House.

So, in our effort to be consistent, we've included that. So, some of the information, most of it, is things is that we would have been asking for, anyway, but there have been some changes and there will probably be still a few more tweaks to make

sure that we're fully in alignment with that data fields collaboration. But I think, for the most part, for this initial round of the online submission, we're pretty close to that.

DR. WINKLER: Just, I'll mention to you that the collaborators in that group, so that we all kind of look at a measure the same way, included CMS, included NCQA, included the Joint Commission, included the PCPI. So, hopefully, you're going to start seeing a standardized way of presenting measures from all of these organizations so, you know, it just isn't disjointed when we are looking at presentations of measures.

DR. PACE: For those of you who have this on your computer, are you able to cursor over and see the comment? Okay.

So, we're having some problem with this particular -- yes.

DR. JEWELL: At least some of these, it's linked to the internet. That's why we don't have internet.

DR. PACE: Well, the links are linked to the internet, but the comments are just part of the Word document. So, you don't have to be hooked to the internet to get the comments.

DR. WINKLER: No, but that's only for the Word version rather than the PDF.

DR. PACE: You were sent the PDF version. A couple of you have the Word version. When you get the official ones we want you to work with, they will be the Word version with all this functionality in it. Okay. We don't want you to work on the one you've got, because it's not complete yet.

If you find that that's not working on your computer, as I said, you can display the comments in another format, so I'll just do this, since we're having trouble with the functionality.

So, here you'll see. So, for high-impact, it gives the exact language from the measure evaluation criteria that Reva

pointed out to you earlier. And, you know, data demonstrating performance gap, that will actually give you some of the examples from the footnotes in the evaluation criteria. So, if you have a question about what you're looking for, that's where you'll find it. We tried to embed most everything that relate to that criteria within the document. So, if you're doing this on your computer, it will be, you know, either cursor over it or do it this way. If you're printing them out, you may choose to print them out this way, though they will be fairly lengthy by the time, you know, all the information is input in here.

Okay. Is there anything in particular that you wanted to look at?

DR. WINKLER: No. I just want everybody to be aware how to use the tools that we're going to be using, and even more granular nuts and bolts. The measures that will go to the TAPs, each one of them will get a copy, but ultimately their collective

conclusions, if you will, will be determined such that it's one version.

The TAP final version, if you will, comes to the steering committee, sort of the same thing, you all get to see that. You all can figure out, you know, discuss how you're going to collectively -- you'll each get your own copy and you can draft your own responses and all that stuff, but ultimately there will be one version.

So, you can see us building it, you know, the TAP fills in their section, the steering committee fills in their section, the staff has filled in their section, and at the end of the day this is what's going to get posted as the final evaluation. This is so that it becomes a cumulative, single document. So, again, we're trying to avoid the flipping pages thing and going between multiple documents so that it tells the story in an ongoing fashion.

We get a lot of feedback, again,

trying to quantify to the degree possible the evaluation. It is difficult. We've certainly looked at a variety of ways of doing it, but the evaluation criteria are what they are, and the evaluation grading scale is meant to try and capture some of the -- it's just not a perfect situation. So, I would guess that a measure that has a whole bunch of C's in it, all complete, you know, is probably going to, you know, do better than a measure that has a whole bunch of -- what are they, N's? None's.

And there are no absolutes in all of this. But that's how we're going to carry this through with you, so we want you to be really familiar with these documents and understand all the things that they can do for you. And, we're also very open to any suggestions you have since you're the first steering committee that's actually going to be working with them. So, help us make it better.

DR. BURSTIN: I'll just add, this

is a little unusual of a project as well, in that we don't often have the TAP chairs oftentimes come and present to the steering committee. We've made you guys members of the steering committee quite intentionally. First of all, we get consistency across all the conditions, and secondly, you get to come and bring that sort of collective voice of the evaluations done.

We'll obviously have the data on the evaluations done within the technical panels, but I think we're hoping it will give more consistency for us across the various technical panels.

DR. PACE: And we're also trying to, in this process, delineate TAP role and steering committee role, and so we really want the TAPs to focus on the specifics of the subcriteria, and you see that, you know, certainly to rate those, and then to have some kind of summary about what are the strengths and weaknesses based on the review of those

subcriteria that then will get fed to the steering committee in terms of their deliberations about whether importance is met, whether scientific acceptability is met, et cetera.

CO-CHAIR DUBOW: So, I have a question about actually the specific process that we should expect, when the next time we meet we will be reviewing measures because they will have gone through at least some of the TAP -- all of them. Okay.

So, do you still assign measures to a primary and secondary reviewer in the steering committee? Can you describe a little bit of what the process will be?

DR. WINKLER: Well, how we do that, how we handle is very much dependent on the volume of measures. When we're dealing with large numbers of measures, like 200, yes, we break it down. When we're dealing with, for instance, the steering committee, and I'm envisioning, to date, all I see in terms of

cross-cutting measures is four.

Okay. You know, we can think about, you know, do we need to break those down, or can everybody, you know, take a look at four. If it were 20, certainly, we would break it down into primary and secondary. It's just a matter of how we divvy up the work and make it reasonable. At this point does anybody see that that's an overwhelming burden that we need to break it up?

Okay. Good. Ten? Yes. Okay. I mean, that's where I start seeing some, so I'm just saying, you know, I don't expect it, I'm just saying --

CO-CHAIR DUBOW: People still have to be familiar with the content of the measures.

DR. WINKLER: Absolutely.

CO-CHAIR DUBOW: You know, you can't not --

DR. WINKLER: Right.

CO-CHAIR DUBOW: You know, even if

we get a lot, you know, even if it's a high volume, we still have to vote as a steering committee, so you can't not know what the measures and their properties are.

DR. WINKLER: Right. Exactly.

CO-CHAIR DUBOW: You're not, you know --

DR. WINKLER: One of the things I think, because, for the cross-cutting measures you don't have a TAP who's going to do some preliminary groundwork and kind of point out the big issues to say, here, look, this may be a land mine. I think that it would be useful -- talk to Joyce about this, is scheduling a conference call for the committee and it will probably be February, you know, January, February, to have a preliminary discussion.

You'll get the measures. You'll have the information. You know what the criteria is. You'll all have a chance to look, and we'll have a talk. Not necessarily decision making, but let's talk about it.

Where do you see them? How do you see the evaluation criteria? What are the questions? Because you may have questions that you need more information for. Fine, give us a chance to go get it.

The measure developers are part of that conversation. You can ask them questions, and that may change. Hearing somebody else's conversation may help how you see things differently.

So, it will be an opportunity for you to work as a work group, if you will, large, to do your initial review before you have to make your decision, and I think that might help you, so you'll act as your own TAP, if you will, so your TAP will meet by call before you come to the final meeting to discuss the cross-cutting. And that might make it just a little bit easier.

The decision making meetings tend to be fairly intense meetings, so you do have to kind of be geared up and ready to work.

And so, we're still uncertain for the number of measures, but even still, we're going to have to do the evaluation. We've got 50 and we know of at least, you know, six more, so we could be talking 25, 30 measures, and we will be talking about complex measures like outcomes. That's a loadful. That's a lot of measures. So for the discussion, and we'll have two days to do it in.

So, you know, getting familiar with the measures, becoming familiar, thinking about them before you come to the meeting, ready to kind of do the final discussions and decision making, it just makes those days go a little bit easier.

They are never easy, but --

CO-CHAIR DUBOW: And before we adjourn today, Alexis is going to tell us about polling us for our dates, because we need to get it on the calendar. Ideally, the entire committee will be here for that spring meeting, whenever it is.

DR. WINKLER: Yes.

CO-CHAIR DUBOW: Because it's really, really hard to do it by telephone. So, I hope that everybody will be able to do it, and we will poll for the best dates for the majority.

Barbara.

DR. YAWN: One of the things that I think might help out sort of a primer, and I don't think that you planned it in your other discussions today, is risk adjustment 101. Risk adjustment is going to be so crucial in these outcome measures, and I just looked down this one and the risk adjustment here is all patient refined DRGs, age and gender.

Well, what does that mean, and how useful is that and what does it take into account, whether it's strengths, whether it's weaknesses, what are the most common risk adjustment mechanisms? And I know there's 15, you know, you can sort of look over the 15 and

say, okay, here are the three most common risk adjustment methodologies used because they aren't all created equal, and I think it is just real important for everyone to understand.

DR. WINKLER: All right. Would you recommend a conference call just with that as the focus of it?

DR. YAWN: That would be what I would recommend, a webinar, a conference call.

DR. WINKLER: Webinar.

DR. YAWN: Yes, I think webinar is much better because then we could have some slides up there in front of us.

CO-CHAIR DUBOW: It would be stored, wouldn't it.

DR. WINKLER: Yes, that's not a bad thing, yes.

CO-CHAIR DUBOW: So if they couldn't make it they could have it.

DR. WINKLER: And then it would be stored.

CO-CHAIR DUBOW: Actually, it would probably have utility for more than just this committee.

DR. YAWN: Yes, because, you know, I'm going to want to share that with my TAP, and whether we try a conference call with just the steering committee and then ask the TAP representatives, say, could I have, you know, whoever gave it support me in giving it or, you know, see one, do one, teach one, or whatever we do.

DR. BURSTIN: I just want to also point out that it's really important, the actual measure evaluation forms, themselves, the measure submission forms themselves have a lot of citations and evidence. So, you know, we could take a 10,000-foot view of risk adjustment and the important considerations like if doctors should present on admission or as close to admission as possible.

But you're still going to get into very condition-specific nuances, for example,

AMI has been, actually interesting, fairly well-validated using these kind of data, whereas you wouldn't necessarily think that for other conditions. So, again, the devil's in the details for some of this but I agree, we should give you some high level of review.

DR. YAWN: Some people have probably never really spent much time thinking about risk adjustment. Process measures tend to have a little bit less with risk adjustment than some of the --

DR. BURSTIN: All of you who were chosen have thought about risk adjustment as well as the TAP.

DR. YAWN: No, the steering committee, some of the TAP people that are consumers, for example, may not have thought as much.

CO-CHAIR DUBOW: Are there any other questions, comments, observations about the measure evaluation form or the process it undergo, because we have one more item to

address before we adjourn for lunch, and then for the day.

DR. PACE: Joyce, can I make one--

CO-CHAIR DUBOW: Please.

DR. PACE: I just want to make one other comment, and I think Reva's mentioned this throughout, but as you see here in this pink area, the steering committee makes, on importance, for example, that's the threshold criterion, and was it met, yes or no for that particular one. But we do ask for you to think about the rationale. So as Reva was saying earlier, why yes or why no, and we are really continuing to push on our reviewers to ground decisions in the criteria.

So, just encourage you to continue to work with us on doing that and we know this is a process and a learning process for all of us as we continue to evolve our processes and criteria, but try to keep that in mind.

MS. GROAH: On the citations, does staff go back and validate those or look at

those or should we be concerned about that?

DR. WINKLER: Actually, I'm hoping that that's something the TAPs will be able to help us out with because, hopefully that's where they've got the expertise. Hopefully, they know that better than us. I don't think we can go back and do all of them at a staff level, but at the TAP level we can certainly say, you know, are these the right ones, is something missing? I think that's the clinical expertise of the TAP members, I think that's one of the important ways of using them.

CO-CHAIR DUBOW: Okay. So, there's one more item, and that is to discuss the gaps, and a way of thinking about addressing the gaps.

DR. WINKLER: I'm not trying to belabor this, but you all have been coming to me with some outstanding ideas, and I just want to be sure I can capture them to the best way possible in terms of approaching the

second goal of the project, which is identifying the gaps in the outcome measures.

My thinking along that, sorry if it's not in projection form, my thinking along that is a lot of people like grids. David's been sitting here, you know, doing this.

Has anybody else written one of these? Okay. If you do, would you share it with me?

What David has done is, he has across the top row are the condition areas, and down the side are the types. And that was why I kind of laboriously took you through those types to be sure what was in was out, embellish them, make them the best they can be, because I wanted to do this. And then, what he envisions is that we put in the measures, the outcome measures that we've already endorsed into little boxes,

DR. HOPKINS: They are the ones that are on the table -

DR. WINKLER: Right. And the

candidates. We can do that, too. And I had sort of thought of that from a, you know, each TAP has got it's own page as opposed to a single one, but whatever, to help identify where those gaps are. Another thing is Pauline brought up a slide that I guess originated with the FDA and patient-reported outcomes where she was looking at data sources, but not in a traditional way, but for instance, where would the data for the information come from. One was the patient. It says patient information or caregiver. The two I think are very similar, they are external to the health care system.

Another would be the clinician, whoever, health care person, provider, representative, observation, and then objective things like the blood pressure, the lab result, something hard and fast nobody's observing or interpreting, as types of data, in terms of where you would get this.

And you know, I was very intrigued

with it because I can see, as we were talking about functioning, you may have measures of function from the patient's perception, and you may have measures of function from the clinician's observation assessment. And those may both be very useful. Maybe not in all of these conditions, maybe some, maybe not. What are we doing?

Yes, unfortunately, how do we share it with everybody? At lunch, come look at Pauline's thing.

But clinician-reported physiological, which is more objective. The caregiver reported or the patient reported and I think the two, it's a proxy report for kids or other folks who need the proxies.

But I just loved seeing this, and it just kind of chinks something. So, those are the sort of things that you guys are just an incredible resource for.

So, David's got his grid. Pauline's got her slide. Barbara, yesterday,

talked a little bit about, you know, breaking down function, the role function, occupational function. You know, I'm going to try to embed all of this. I also want to use the -- so it's a care framework to help us thinking about over time, outcomes over time, because addressing Linda's issue, you know.

So, the big, final end point that's down the road some ways, all important, but in a measurement world, sometimes very challenging. What are the more short-term outcomes? How do we think about the different processes as a patient goes through an episode of care, you know, yes, hospitalization may be, you know, for an AMI, but there's the post-acute, and then there's the secondary prevention, and then there's all the impact of that disease.

Those that are on a, you know, I'm going to do okay trajectory, versus those that are on a not doing so well trajectory, what are the outcomes of the care that person is

experiencing that you'd want to have information about. So, these are the kinds of characteristics of ways of framing the question of where are the gaps in outcome measurement that we want to identify to address the second goal of the project.

So, my question to you all is: What other good ideas have you got brewing there, because I know they're out there. You're starting to kind of share them, but I want to try and take advantage of the fact that we're here together today.

If there's anything out there brewing, just as Pauline did and David, I mean, you didn't share it, I took it. But anything else then, as well as if, on your travels in the next couple of days, as you sort of mull over and think about the conversations we've been having, for the last two days, thoughts on how we might portray this, to be able to do the analysis of where are the gaps in the outcome measures, what

outcome measures would really be useful to a wide variety of audiences that would help this whole process?

How do we fill those gaps, what do they look like, because I think for different topics and different conditions they're going to be different. Certain topics are going to lend themselves to certain types of outcomes more than others. And that may be, you know, some of the acute and chronic, some of the, you know, natural history of the disease, the expectations, what we know about the efficacy of treatment, all sorts of things.

So, any way to help characterize it at this point is a good idea. We're trying to figure out the best way to move this one forward. And so, this is sort of another assignment, if you will, for goal two of the project, is if you've got any additional ways of thinking about it that you'd like to slice and dice this, please share them.

Your first assignment, of course,

on goal one was, are there any measures out there, do you know any measures out there, and get them to us, as well as beginning to orient yourself around the measure evaluation criteria as we go forward. So, I'm kind of open. At this point we've pretty much reached the end of our agenda. Lunch should be ready. Yes. And so, I want to open it to any questions that you may have. Let's give this a final opportunity to talk about the things we've talked about and ask questions.

DR. HOPKINS: So, goal one is to hustle and get measures submitted. Is there any chance that you guys can push the deadline on that? It's like two weeks from now, or one week. And, you know, this is the first opportunity I've had to really think more about the gaps and the great opportunity we have.

DR. WINKLER: Yes.

DR. HOPKINS: To fill some of those gaps. But I'm really going to be hard-

pressed to see how we can identify the source and get them to fill out the form and all that by October 31st.

DR. WINKLER: No. At this point that's an open call for -- it was just part of it. But in terms of getting the measures like, you know, the desperation aspect of it, we'll take them. We'll figure it out.

DR. HOPKINS: Okay.

DR. WINKLER: Okay. Just let us know, we'll work with you. It's not a problem. Don't consider that a limiting factor. We'll deal with it.

Now, it will be very hard if you come up with them in April. Okay.

DR. HOPKINS: A month or two.

DR. WINKLER: Yes. If we really would like to see things, you know, no later than the end of November.

DR. HOPKINS: All right.

DR. WINKLER: I mean, we can probably still put things in in November.

Beyond that, it's going to get a little bit tough on some of the topic areas.

DR. BURSTIN: Well, some of that depends on the dates of the TAP.

DR. WINKLER: Right.

DR. BURSTIN: That we're conceiving to be in December, so we need to leave them sufficient time to review the measures and not just dump it on them right before the meeting.

DR. WINKLER: Right. Exactly. Right. And that's why I'm saying.

DR. KEALEY: So I just need a little clarity about the previously NQF-endorsed outcomes measures that you sent us. What exactly is the relationship between those and these new measures we're getting? Are we evaluating those as well or it's just --

DR. WINKLER: No. Those provide the context of what you're doing because the work you're doing is to add to that portfolio. One of the things we're likely to do with

David's grid is make a grid and populate it with those, find out where we do have measures in italics or pink or some other color.

We'll put in the candidates, see how it fills out the grid, and then we'll look at the empty spots. So it's an ongoing building of a portfolio of outcome measures.

It would be very important for you to not look at the candidate measures without the context because if you saw the one AMI mortality, we've been through AMI mortality before.

There are endorsed measures around AMI mortality. It would not be appropriate for you to evaluate that without considering what's already endorsed, you know, and looking at the big picture, if you will, because our goal isn't to just keep endorsing multiple versions of the same measure, but how do these all fit, does this bring something new to the table, does it -- is it better, you know, is it a better mouse trap, is it -- and so you

need the context of what other measures are out there to build this out.

So that's what it's for. It's the context.

CO-CHAIR DUBOW: And don't forget, otherwise, the measures, the endorsed measures are routinely -- they are maintained, and they are reviewed every three years unless there's a reason to do it more frequently.

So that process happens, but if there's a measure coming in anew that can be compared, we're looking for best in class.

DR. KEALEY: Okay. So, yes, so the criteria, the four criteria, you said, have changed in the last year, but because these are renewed every three years, we can assume that even if they predated this reclassification that they are still valid and they are going to be looked at?

DR. WINKLER: Yes, and the other issue is, as you are doing your comparison it will be difficult for you not to get into some

of the details of them, and if you see issues with some of the currently-endorsed measures, we will collect that feedback and feed it into the maintenance process.

So, you know, every opportunity to really understand what are the best measures, we'll try and take advantage of it.

DR. BURSTIN: We'll actually try to build in for you the date of the next maintenance so you have a sense of how stale or fresh they are and whether you -- you could really make a pretty significant impact on that maintenance review by giving us input as to the existing measures in addition to the ones that come to you.

DR. JEWELL: So specific to that conversation, at least some of the outcome measures that you have in that file were originally endorsed as time-limited to begin with, so their window is shorter. It's two years.

DR. WINKLER: Yes, it's happening

right now.

DR. JEWELL: And part of the time-limitedness or, at least my memory is that a big reason for the time-limitedness was because they really had not been tested for the purpose of quality evaluation, so I think when that information comes back around, these newer criteria that have come into play can be applied at that point in time.

DR. WINKLER: Right.

DR. JEWELL: Regarding the gaps question, I think it's going to be important for us to be clear about whether we think every type of measure on the grid needs to be filled.

DR. WINKLER: Right.

DR. JEWELL: And I know you didn't say that, but I think when we're talking amongst our TAPs and others, we don't want to confuse gap with every little block in a grid should be filled with a measure for a condition -- this type of, you know, every

type of measure out there in the condition. So I just want to make sure that we agree that that's true.

And the third thing I just wanted to bring up was, it seems to me the biggest gap is that there are lots of outcome measures out there already, but they weren't, again, designed with the aggregate in mind. And so really what we're talking, at least in my world and in Anne's world, we are really talking about measures that have potential to be aggregated but just hasn't been thought of that way and so that's really where the gaps, I think, may come for others as well.

So it's not that the measures don't exist.

DR. WINKLER: Right.

DR. JEWELL: It's just that they haven't been thought of in that particular framework.

DR. PACE: And I think that's good information to know, whether it's a patient-

level measure, but it hasn't been developed into a provider-level quality measure, so that would be useful information because that's a great building block.

DR. WINKLER: Right. Exactly. Yes, I think because this is doing the gaps analysis part of it is such an important part of this project, we're going to be able to look at the nuances around that and include that.

And absolutely, Dianna, if I didn't emphasize, yes -- not all of those types of measures will be appropriate for all of the types of -- and where it's not appropriate, we'll just say so. You know, it's just not -- you know, not a particularly useful outcome for that particular condition, and that's part of the assessment and part of the analysis.

We wouldn't want somebody to go create something that is meaningless. So I definitely agree, and thank you for making it

explicit.

DR. BURSTIN: And just to add to that, I think, you know, for example, I'm thinking of Gallo, you know, we'll have no wine before it's time, I think there's also sort of a sense that although there's a real sense of urgency here -- I know it's right.

Go stand with the drunk under the street lamp was just too much to, you know -- all these street lamps down the road here. But too many bad analogies today.

But I think that, you know, there are gaps that are going to be identified clearly. Some can be filled in the time course of this project, and some can't.

And so I think the idea of saying there are some that really could be created into a quality measure, the idea that that's going to happen in a month or two in a high-quality way is unlikely.

So I think it's just as important that we identify what needs to happen, even if

it doesn't happen in this current project, as I mentioned, we now have the resources that I think we should be able to go back and say, in a year, let's reopen the outcomes project and bring back in those measures that were identified as gap areas and bring them back in.

I know there's a sense of urgency. Let's get this first set done. But I also just don't want people to feel like we have to sort of push so hard that things are coming in that you're just not comfortable with that won't make it through the process.

DR. PACE: I'd just like to make one comment about the evaluation criteria. We talked about them being revised last year, but I do want to mention that, in essence, they are the same. I mean, NQF has always had criteria about importance, scientific acceptability, usability, and feasibility, even to the extent of, you know, reliability and validity being under scientific

acceptability.

So there's more clarification, there's more detail and guidance, but I just want to make sure that we're understanding, it's not like a totally new ball game. I mean, these have always been kind of the expectations, but I think it would be fair to say we're ramping up and trying to expect more of meeting those criteria in more rigorous ways and will continue to make that evolution.

CO-CHAIR DUBOW: Any other comments?

DR. KEALEY: Yes. I was wondering if you guys could walk me through. I know we talked a little yesterday about, say, the unintended consequences scenario.

So we endorse a measure; CMS puts it out and starts using it; unintended consequences happen. From the end user, what do they do, how do they effect change, and kind of what has been the experience with the time line between users starting to have

trouble and boom, boom, boom, it goes up the chain and back down the chain, and that measure's gone.

DR. BURSTIN: To be honest, it hasn't happened a whole lot so, you know, I'm giving you ns of two or something like that. It's been very, very small.

And our experience has been when there actually has been evidence, like the pneumonia example I gave you yesterday, of significant unintended consequences, we had that ad hoc committee impaneled within a couple of weeks of publication. The measure was revised and brought to the board, I think, within a month or two. I mean, it was very rapid, and CMS adopted the new measure.

So I think when there's truly evidence, and that's the biggest piece of this, when there's evidence. And the problem is we've had other discussions, for example, about perceptions of unintended consequences, a whole discussion around the 30-day mortality

measures and this question of the fact that you couldn't exclude patients who were put onto the hospice benefit beyond day one.

And this was a huge issue that came up but, you know, the evidence for the unintended consequences wasn't really there, although I think there was a lot of perceptions of that.

So I think that's one of our challenges, and which is why we talked a lot about making that robust feedback loop stronger, but I think the key for us is we need to hear from people when there are measures with unintended consequences, and we just don't hear very much, but I think CMS is responsive.

DR. AMARASINGHAM: But I guess, in that scenario there was a peer review publication that needed to occur, right? So that's probably nine months, nine to twelve months.

DR. KEALEY: Right. For people to

start looking around, noticing issues coming up, critical mass, get together somebody to say, hey, I'm going to study this, publish it, then it comes to NQF.

DR. BURSTIN: And it doesn't -- and I should clarify that that -- that it doesn't, from our point of view, require a peer reviewed publication. We all know how long that takes.

Another example we've got going on right now is there is some debate within the surgical community about hair removal, a measure I spend way too much of my time on in an extraordinary kind of way.

But a whole issue of whether it's actually, you know, you're not supposed to shave, you're supposed to use depilatories or other mechanisms, but there's some issue about whether it's actually appropriate for neurosurgery.

So we don't require a huge number of, you know, publications to say this is an

issue. There's been some concerns from the field, and so we're convening an ad hoc maintenance review committee to look at the evidence as it exists.

But, again, we can only do that if we know there's a problem.

DR. KEALEY: And so if end users who I doubt fully understand kind of the way these measures go through the system, so if they're reporting back to Medicare or complaining to their local Medicare person, they know to move it to NQF or do they mull on it a while or what's to ensure that the word is getting back here?

DR. WINKLER: At this point, absolutely nothing, except it's sort of a random thing, which is why we're trying to get the word out to you all and the people you work with, that bring those to us as well as to CMS.

I would have to say there's probably no guarantee that that communication

occurs. Sometimes it does, but I think sometimes it doesn't. So if we're talking about measures that we've endorsed, we really want to hear about it, and we're happy to hear about it, you know, sooner rather than later, so that we can keep an eye on what's going on. That would be my best recommendation.

DR. PACE: Certainly if we're talking about end users and unintended consequences to providers, professional associations also present an avenue for getting information back to NQF, which is, I would say, how our members typically do it, that they are not as willing to rely on fiscal intermediaries of any kind.

CO-CHAIR DUBOW: But the point is that they need to contact NQF so that it's, you know, it's assured that NQF knows about it so that NQF can take steps.

I mean, if the issue is to address the endorsement, either to reaffirm it or to withdraw endorsement, NQF has to initiate a

process, and so the feedback about how measures are doing is necessary.

DR. AMARASINGHAM: I was just going to say, because I think Burke's point is very important, because I think a lot of the clinicians and others may not know the rigor that it's gone through, and then think that it may take a long time.

The question is whether NQF should actually, rather than just be a purely reactive process, should have a proactive process about measure surveillance, exactly how well is the measure working, you know, look at quality assurance with respect to the data sets that's coming back for it.

Because if it's all in the end user's -- just thinking about how even a hospital is putting an EMR together, clinicians don't report problems that they have with EMRs. I can't imagine they're going to do anything with measures. So just those kind of considerations.

DR. BURSTIN: Those are all really important points. For the first time we're actually going to be doing a formal, external assessment of the impact of NQF-endorsed measures to give us a better sense of how do we even begin to -- I mean, I have to be honest. I wouldn't even know how to begin tracking some of this without being reactive, but actually active surveillance of some of these. And so we're hoping that this work that will be done externally will help us sort of think through some of the paths to getting at that.

But I agree completely. And I think Dianne's point is well-taken. NQF is an organization of organizations, and so going through your professional organizations or consumer organizations is probably the best mechanism.

DR. KEALEY: So does NQF track or do they ask to be notified if somebody's using one of their measures? Do we have a database

of who's out there using the recommendations?

DR. BURSTIN: No, and this is one of those interesting points as well. We've talked about this a lot as well because we are not the measure developer.

We are not the measure steward. So our hope is the measure steward should know that. But, again, we're trying to think about, you know, where is out logical fit in that measure steward, measure user, endorser kind of loop, and advice and thoughts about that are very welcome.

CO-CHAIR DUBOW: There is some activity going on at the AQA for the ambulatory measures to make some kind of an assessment about reports that are out there and which measures are being used, but I don't know what the -- it hasn't fielded yet. It's in the, you know, it's in the development stage.

And I don't know what will come of it or what the response rate will be, but in

the past AQA has done some surveying of the health plans to see which measures they are using.

So, you know, there are these pockets of inquiry to find out. But, you know, I mean that was done a while ago. I don't know if that was the point where AQA was actually using only -- now AQA endorse -- uses endorsed NQF measures or supports using NQF measures.

So I don't know what the status of that was, but --

DR. WINKLER: Yes, and one thing I would just add is, part of the maintenance review is one of the questions, one of the most important questions is, is the measure in use and how is it being used.

And the idea of keeping, collecting that in a database that becomes available, that we can use, and then even encouraging people to let us know prior to the maintenance, you know, is something that I

think we can certainly consider going forward because I can, you know, perceive the utility of it.

But we will have a foundational way of filling initially the database with our maintenance information, and that's what we're starting to collect this year. So, you know, we do have the beginnings of something.

Before it was a completely random thing. Who did you talk to, who did you hear from and what did you trip over, as opposed to any systematic way of collecting the information.

But I can see that we kind of have the beginnings of something that we could certainly work on and it's a great idea.

DR. KEALEY: Yes, I mean, I guess the concern I would have is you've got people out there using these measures. You're updating them using new science and everything, and how do you get the word out to all the people who are actually using them

that, oh, that measure we endorsed three years ago is not any good any more and we think this one's better, and right now you have no way of getting that out there.

DR. BURSTIN: I guess it would all depend as we improve our online database. I mean, at this point you can at least see what's endorsed or not endorsed. Hopefully, you'll be able to, in fact, track the timing of a measure when it was last endorsed, when it's up for maintenance, did it make it through maintenance.

And currently, just so you know, the first set of 40-some-odd measures that are going through our maintenance process are now on the NQF website and posted for public comment. So we're hoping to actually do a more proactive polling of what's people's experience.

CO-CHAIR DUBOW: But it does suggest that the implementers need to be knowledgeable about referring to the NQF

database, now that it is maintained. You know, now you actually have access to the current information.

So it behooves an implementer to look carefully at what, you know, the status of the measure is.

Dianne.

DR. JEWELL: So I haven't been a participant in the maintenance process. I know that you ask how's it being used. I heard the two questions. Is there a specific question about unintended consequences because I'm thinking like adverse event reporting in research trials, you know, in clinical trials.

DR. WINKLER: Essentially, and Helen, help me out here, but what we're asking them to do is take the original submission criteria and ask where it changed. And unintended consequences falls into one of those categories.

So those are the kinds of things that we're looking for, what did you learn

about the measure's behavior, both good and bad, as well as how it's being used.

DR. BURSTIN: The only thing I would add to that as well is we've had some discussions actually as recently as last week with our board about what are the requirements of maintenance in terms of public reporting and use of the measure.

So should a measure continue to be, you know, endorsed by NQF if no one is using it. And I think we're still trying to figure out exactly what that means.

But if nothing else, I think we are continuing to raise the bar in saying, okay, it's been endorsed for three years. As best as we can tell, no one's used it.

And actually, the secondary question is not just is anybody using it, are they using it in public reporting, but if you've used it, does it actually help you improve quality, I mean, the QI piece as well.

So this is definitely a work-in-

progress. We finally have the resources to really be able to do maintenance in a way that we've never been able to do before.

I think, if I had to predict, I think the portfolio would be half the size it is in a couple of years, which I think probably would be right-sizing it to where it should be.

DR. YAWN: Do you also have some funding to look at people who currently don't use but do public reporting? And Minnesota is one of the examples I always use because we have a public reporting process.

They make up their own measures on a regular basis, and there's also ICSI who takes measures and then redoes them. And so I'm fascinated, and there's probably other states. You ask them why and try to find out.

So I'll look forward to that kind of information because I think that's crucial as NQF becomes recognized as the resource of endorsement and why --

DR. WINKLER: Just in response, are you talking about Minnesota Community Measurement and ICSI? Actually, we evaluate their measures and, you know, we use --

DR. YAWN: Yes, but they use them before you ever have time.

DR. WINKLER: Absolutely.

DR. YAWN: Believe me. I know because they sort of say, "You've got to finish looking at this today because we're going to start tomorrow." And so --

DR. KEALEY: But isn't that what we are asking? We want people to kind of use these and try them and give us good evidence.

DR. YAWN: Yes, but they change them every year.

DR. KEALEY: I know. I live there. I know.

CO-CHAIR DUBOW: Okay. This has been a fruitful discussion, I hope, and if there are no other questions, I think I see a couple of new faces in the audience.

If there's any public comment -- and also to ask the operator if there's anybody on the phone who has a comment.

OPERATOR: All lines are now open.

CO-CHAIR DUBOW: Okay. Anybody here?

DR. WINKLER: No. Alexis --

CO-CHAIR DUBOW: Okay. So we have Alexis to talk about getting our act together.

MS. FORMAN: Just quickly, this is a tentative time line. We're still waiting for our approval from Health and Human Services, but we had to come up with some dates, and so we would like to start the TAP meetings for phase one in December.

And I will work with the TAP chairs to make sure that they can attend the meeting. So if they aren't available on this particular date, we can always change the date. We just needed something down for them to approve. So it is flexible.

(Off-mic comment.)

MS. FORMAN: Yes. These are one-day face-to-face meetings, and it will be in D.C.

So, phase one, we have, starting in December, the beginning of December, and we've only scheduled two, and I think we're going to change cardiovascular and we're going to move that date to December because we do have a lot of measures under that TAP, and we'll be doing some measure maintenance possibly.

So we'll have multiple conference calls, probably, with that TAP. And any of the conference calls that we might need to have after this in-person meeting, we'll send out a survey so we can make sure that everyone could attend that call. So we'll do an availability survey.

And for the main steering committee, we want to meet towards the end of April, but, again, this will depend on your schedules, so, again, we'll do an availability

survey, and we'll survey the entire steering committee to make sure everyone can attend.

So, hopefully, this should get approved within the next one to two weeks, and then you'll be hearing from me, especially the TAP chairs, to finalize the dates of the in-person meeting, and then we'll get the dates for the second steering committee meeting, and any conference calls that we have in between we'll also do a poll so that we can make sure the majority can attend.

CO-CHAIR DUBOW: Alexis, can you please let us know when those calls happen, just so people can put it on their calendar as a reference. I think the most critical issue, though, is to get on our calendar the date this next meeting because as I said before, it's really important to try to be here, and it's going to be a two-day meeting, is that correct?

MS. FORMAN: Yes.

CO-CHAIR DUBOW: Okay. Right.

We're going to have a lot of work. So, but we'll have some preliminary stuff. We'll have the opportunity to look at the cross-cutting measures, and we might as well get those dates on the calendar as well, and even this tutorial on risk adjustment that Barbara suggested sounded like a really good idea.

So, even though we're not meeting until April in person, we will have multiple opportunities to be thinking about the activities related to this steering committee.

MS. HAUGEN: And just to clarify, it's the 28th and 29th? That's what you're targeting, April 28th and 29th?

CO-CHAIR DUBOW: No. No, we're going to poll for those dates.

MS. FORMAN: Well, wait. It's just tentative. We had to put something down, but we'll poll.

(Off-mic comment.)

DR. WINKLER: Those are ball park. Consider it ball park.

CO-CHAIR DUBOW: You'll be polled for the dates. These are just sort of to provide some --

DR. WINKLER: The month is correct.

CO-CHAIR DUBOW: The month is correct.

MS. FORMAN: Yes. The week, it could be correct. It depends. We're going to work with the TAP chairs, but we had to put something down within a day or two, so we had to give them like a skeleton type of time line.

CO-CHAIR DUBOW: We're just going to have to be flexible until we're polled, and then respond as soon as you can so we can firm up these dates and put them on our calendars.

MS. FORMAN: The polling will occur once the Department of Health and Human Services approves it. So we have to make sure because we're under contract with them. We have to make sure that they are okay with our

time line.

CO-CHAIR DUBOW: It's within a week.

MS. FORMAN: Yes, it should be within a week or two.

CO-CHAIR DUBOW: Yes.

MS. FORMAN: Because they've had this time line, so it's just making sure they're okay with it, and I mean, we apologize, but we have to have them approve it before we can schedule dates.

CO-CHAIR DUBOW: It's the nature of working with a contract with the government. You know, we just have to be flexible until they say okay, and then we can get into action.

DR. YAWN: Alexis, when we poll the TAP committees, I think it's very important that we let them know this committee meeting will begin at 7:30 a.m.

The reason for that is, you know what happens, people want to fly in that

morning. They get there at ten, and then they want to leave at two.

And so I think we have to make it very, very clear that we're going to start really early in the morning, even if we don't, so we can get them there for the full day.

CO-CHAIR DUBOW: Ruben, did you have a comment?

DR. AMARASINGHAM: Just a quick question about time. So then after our meeting in the spring, is this committee continuing to work through October? So it's for another year after that or just --

DR. WINKLER: In reality, the role of the steering committee, the biggest part of your work is through the meeting in April. After that you do have several activities. We'll go out for public comment, and you'll come back to respond to those comments.

Then it will go to voting and to CSAC. You may or may not be asked to have any feedback or responses. And then it will

ultimately be endorsed so that, pretty much, that will be the end of the steering committee's real work.

Occasionally we've had situations where we've come back to you for a question. We said, hey, can we get you guys together on a conference call, and something has come up. But those are very unpredictable and tend to be sort of on an ad hoc basis.

So the vast majority of your work will be done by the summer, though there could be an occasional, hey, you know, we want to check in with you on something after that time.

MS. FORMAN: And then, again, when we do set up the conference call to review the comments, we will, again, poll you to make sure that you're available.

CO-CHAIR DUBOW: Okay. So, you know, the idea of polling obviously is to get as many people on the call as possible, so you know, we need to be flexible, but the NQF

staff will accommodate us to the extent that they can.

So I think with that --

DR. WINKLER: I just want to mention lunch is out in the hall. Please, you know, we bought you lunch, please enjoy it.

CO-CHAIR DUBOW: We actually finished early because we were really efficient. I hope that this was a productive meeting and that you have a good sense of what the expectations are of us.

We have a lot of work to do between now and April, but the staff has even more work than we have, and don't forget to see if you can find good measures that fit the scope of the project and to think about a framework to add onto David's really neat grid. And I'm sure people will just hang out here for lunch, but I wish everybody who's traveling good trips, safe travel, and I'm sure we'll be in touch soon.

DR. BURSTIN: I just want to add

my thanks to the steering committee and Joyce and especially to our staff who have worked, obviously, very hard to get all this stuff together.

CO-CHAIR DUBOW: I was going to say I wanted to thank the staff, too.

DR. BURSTIN: And in particular I'd like to --

DR. WINKLER: Don't be strangers. We work for you.

DR. BURSTIN: Let's hope there's no hurricane in Cabo San Lucas for Melissa and Alexis's part for next week when they're supposed to be on vacation.

(Whereupon, the above-entitled matter was concluded at 11:59 a.m.)