NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
July 17, 2012
12:00 p.m. ET

(Lorelei Gorean): Hi, everybody. This is (Lorelei Gorean) from NQF. Thanks very much for joining the GI/GU call today.

I think before we get started, I'll say I know Taroon is on the phone, senior director Ash, data project manager, is also on the phone and Evan, our analyst is on the phone and perhaps, we could just open up the call to see who else we have on the call.

Robert Ellis: Robert Ellis is here with Consumers' CHECKBOOK.

(Lorelei Gorean): Great. Thanks, Robert.

Johannes Koch: Johannes Koch from Virginia Mason.

(Lorelei Gorean): Great.

Any other steering committee members?

Andy Baskin: Yes. Hi. It's Andy Baskin.

(Lorelei Gorean): Hi, Andy.

Andy Baskin: And I think I'm doing this for the second time. I think I did this one about a year – a year ago. I'm trying to learn more.

Taroon Amin: I feel that and I know that a number of people are dialing and listening to the webinar so welcome to you all. The purpose of today's call, my name is Taroon Amin by the way. I'm the senior director supporting this project.

The purpose of today's call is to provide a little bit more of an in-depth discussion around the criteria that you'll be using to evaluate concepts that are submitted to this project. Many of your participated in the orientation call in which we gave you a high level overview of the specific criteria that you'll be using to evaluate.

This call will be – is structured to be a little bit more of an informal discussion. What we will do is – at any point in time, feel free to stop me to ask any clarifying question. Any comments that you may have will be welcomed at any point. What we'll do is we'll also walk through a number of the materials that we provided, specifically the steering committee guidebook, which we have created and are piloting as part of this project in order to help you have a one-stop resource for any questions that you may have as your evaluating measure in this project.

So I'll actually begin by, as you could see through the screen share, this is your SharePoint site. Hopefully, you had an opportunity to log in and begin to use some of the functionality here. As Evan pointed out to me earlier today, there are – as far as July 11 call goes – as far as the July 11 call goes, there are two new items here in the July 11 call, which include the recording and the transcription in case you're not actually able to join us for the orientation call.

As you scroll down here on the left side, you'll see July 17, 2012 and here, you'll find the materials that we'll use for today's call, specifically the steering committee guidebook and I will encourage you to – I will encourage you to open that up and follow along.

Male: Taroon, can I interrupt you for a second?

Taroon Amin: Yes.

Male: Well on the main SharePoint page is also the archived webcast of the orientation. So if you want to watch the slides and the recording at the same time, that's located on the main page of the project link.

Taroon Amin: OK, great!

I just wanted to – just a quick question because I'm getting some errors (inaudible) menu. Ash, are you able to see the screen share that I'm sharing here or Ashlie or to (Lorelei)?

Female: Yes, I can see it.

Male: Yes.

Taroon Amin: OK. Great. All right. I'll just ignore the warning signs here.

OK. So in this document here, just to orient you. I know Evan had oriented the committee to this document and we're hoping that this document is useful. If there are any questions or any enhancements that you recommend as we serve in launching this guidebook, for the steering committee members, please send them our way.

To orient you to this guidebook, you'll see a, you know, what is NQF, a description of our mission, governance, the NQF members, and then the description of why NQF endorsed measures are important – the description here, number three of the NQF consensus development process, which gives you a high level overview of the process we'll be using in this project. There's also a description of other strategic partnerships, which are part of NQF, which include the MAP, which is the National Priorities Partnership and I mean NPP and the MAP which is the Measure Application Partnership.

What I'll ask you to focus on actually here is Roman numeral III, which begins on 19, which describes the steering committee – what do I need to know and I will actually move down here to – under Roman Numeral II, Subheading three, Evaluating Candidate Measures. This gives you a high-level overview of the various criteria that we'll be using through stage one and stage two evaluation.

I'm going to actually move to page 25 of this PDF. As you can see, there's a page entry – I worry just you know, for those of you that are not as accustomed to Adobe, there's a page number listed up here, page 25 and I'll continue to toggle back and forth through various different portions of this PDF.

So measure developers will submit concepts to this project and the concepts will be evaluated against the Importance Criteria. The Importance Criteria include three subcriteria and so the Importance Criteria to the extent in which the measure focus is evidence based. It's important to make some significant gains to health care quality and improving health care outcomes for a specific high impact of health care.

Impact is evaluated by three subcriteria. There's three subcriteria – the first being high impact, which you'll evaluate high, moderate, and insufficient. This will essentially as the question of whether the measure focus addresses the high – a high impact area or National Health Care Gold defined by the MPP or HHS.

Secondly, you'll be looking at the performance gap – whether there's actually demonstration of a quality problem and an opportunity for improvement. So in the case, we're actually looking for the measure developers to demonstrate that there is data demonstrating considerable variation and performance across measured entities or that there is an overall less than optimal performance for all measured entities.

So for our subcriteria 1-A and for 1-B, you'll be using the generic rating scale, which as you could see by this hyperlink, which will be – which is labeled definition, table five – I'll just click on that, which brings us to page 33 of this PDF. Table Five describes the Generic Rating Scales for subcriteria 1A and 1B, which will be high, moderate, low, and insufficient in which high is defined as baseline of the information submitted. There's high confidence or certainty that the criteria has been met.

Again with all these criteria, there's a level of clinical and methodological expertise that we will ask the steering committee to use in order to actually

rate high, moderate, low, and insufficient. I think as I describe high impact, the performance gaps are fairly straightforward. I'll spend a little bit more time on the evidence to support the measure focus. The effort – I mean the evidence to support the measure focus is defined by three basic pieces of information that we will be asking for from the measure developers. That includes measure developers providing us the quantity of evidence supporting the measure focus, the quality of the measures of the evidence supporting the measure focus, and the consistency of the evidence and its relationships to the measure focus.

And I will go into a little bit more discussion around the high, moderate, low, and insufficient for these three various sections and this will bring us to page 29 of the PDF, which is table two. Please let me know if I'm going too fast and moving around the PDF. Again, this is more to orient you to this information. You have all these information on the SharePoint site for further reference.

But as we look at the definition here – when we're looking at the definition of quantity, is we're looking for the total number of studies, not simply articles or papers demonstrating the evidence to support the measure focus.

For quality, we're looking for the certainty or confidence in the estimates of benefits and harms of the patient across studies in the body of evidence. We're looking for a description of the study flaws and design and the directness to the actual measure focus in terms of the population and the intervention.

And finally with consistency, we're looking at – we're looking for stability in both the direction and the magnitude of the benefits and harms for patient across the body of evidence. So the specific way that quantity, quality, and consistency is rated, if you're looking at high rating, we're looking for five plus studies with the quality of evidence looking at randomized controlled trials that provide direct evidence to the measure focus and the consistency of that – there's clear and consistent in terms of the magnitude and direction of the evidence base that's supporting the measure focus.

When you move down to moderate, we're looking at two to four studies that support the measure focus, RCTs without serious flaws, and nonrandomized controlled trials that actually have controls for confounders. And we're still looking for estimates of clinically and practically meaningful benefits and harms to patients that are consistent in direction but maybe not necessarily in magnitude.

And finally, what would be a low rating for looking at the quality, quantity, and consistency of the evidence is there are zero to one study with the quality of questionable studies that's either a questionable study design or biases that may implement the point estimate or without adequate controls for confounders.

And finally, when we're looking at consistencies that there is wide variation in what – with the magnitude and whether there is actually positive or negative effect of the measure focus on the incentive population. So yes, that basically describes the body of evidence that we're looking for in this criteria. And that gives you a high-level overview of the three various different subcriteria you'll be evaluating when you're looking at impact opportunity and evidence.

I'll stop there with any questions. I'll – the second half of this discussion will actually go through the steering committee evaluation guide, which will start to give you some examples of the types of information that measure developers will submit to adequately assess these criteria. So if there are any questions, happy to take them at this point.

OK, silence but hopefully, if there are any questions, please feel free to ask them at any point in this call. It is meant to be served informal question and answer.

So if you scroll down on the left side, you'll see Roman numeral III, which is the Steering Committee Evaluation Guide. This Steering Committee Evaluation Guide is intended to put together information as example. Again, the information that is presented here is not necessarily examples of high ratings but more or less provided for illustration purposes only. So as you

scroll down here, this provides you hyperlinks at the various reports that will help you as we're looking at these various different criteria.

I'll specifically point you to the description of the measure concept. This will be the type of information that will be presented in this project. Again, since this is a unique project in a sense that we are going to be piloting a two-stage process which we'll be evaluating concepts for and then fully specified measures later next year or early next year, I should say. So this specific information that was presented for measure concepts include the measure title, a brief description, enumerator statement, preliminary enumerator details without coding information, denominator, exclusions, and a proposed risk adjustment methodology along with mapping to our levels of analysis and data source and setting of care to our – mapping to our taxonomy.

So as you scroll down here, I'm now on page 57, this provides basically three examples, which I'll point to, which is essentially this example of high impact and then we'll provide some two examples of the evidence that we would expect from measure developers.

I will now scroll down to the first example here, which begins on page 59. Essentially, there will be a dropdown menu for 1A.1, which as demonstrates the high-impact aspect of care and we would ask that the measure developers provide information around how basically – whether it affects large numbers of people or patients and a large societal consequences for poor quality of care. But the critical question that we're asking here is around 1.1A.3 which is that we're asking developers to provide epidemiologic or resource use data that demonstrates that the measure addresses a high-impact aspects of care.

And here, what we're looking for is a description of the quantitative data, the number of individuals and the percentage of individuals affected by this measure focus. And we're not simply looking for conclusionary statements such as this measure focus addresses a high impact area of care without any qualifying and quantitative data.

Specifically, we are asked that the measure developers submit information that's related to the underlying and that since actually describing of the

measures specified for the Medicare population that the information actually relates to the burden of illness in that specific population and not speaking broadly about the condition under evaluation. One A4 asked for the citations demonstrating high-impact area.

One – and now we move onto Performance Gap Opportunity for Improvement. Here we asked for the measure developers to provide a rationale for the measure, explaining -- briefly explaining the benefits envisioned by the measure and here what we're looking for in 1.B2 is really data demonstrating the performance gap or the opportunity for improvement.

Again, what we're really looking for here is quantitative data on the actual percentage or the number of patients on the distribution of performance across patients. It should be refined to just the measure focus and the target population. Here, one bit of nuance I'll point out is for new concepts or new measures that are submitted to the project. We would expect that data would be – could be derived from literature, studies, or testing but for those measures that are up for endorsement maintenance; i.e. those measures that were previously endorsed, we would expect that the data actually demonstrates – that it actually shows data from the performance score – I'm using this actual measure as specified for the level of analysis.

So we would actually expect to see the data from using this measure since this has been actually endorsed and unused. And here and one before, we're looking for data on disparities if the data is actually available. Again, we're looking here for quantitative data in order to really be able to assess the difference in disparity.

I'll actually skip over this example around health outcomes because for this project, we will not actually have any health outcome measures that we – we did not have any outcome measures that were submitted to this project.

So I'll skip here to page 71, which gives an example of evidence for process measure. And again, I'll just remind everybody that this information is presented for illustration purposes only and is not meant to describe measures that are – that would meet this criteria with a high rating. So here, what we're

looking to understand is the evidence that supports the measure and what we would expect is that – sorry here. So criteria 1C.2.1, these measure submission forms, you would expect this to be empty because this is looking at health outcomes. So you actually just scroll down here and what you'll end up seeing here is – what we ask for is a brief description or diagram of the causal pathway that actually demonstrates the proximity to a higher – to the desired health outcome.

We're not simply looking for any process measures but we're looking for those that are able to demonstrate true evidence that they are proximal to the desired outcomes. So through 1C.3, we're asking measure developers to submit the causal pathway that describes how the measure focus will actually influence patient-centered health outcomes that are important to patients. And we expect here that this will be a causal pathway, not simply general statements that the measure focus will actually improve the quality of patient care.

What we're looking for here is the description and essentially, when we're looking for evidence, there are three main avenues that measure developers can use to demonstrate evidence. The first is to cite evidence that supported true guideline. I'll make a note here that guidelines are not enough to demonstrate evidence to support a measure. We actually expect that the measure developers will supply evidence – focus us on the evidence that's supported through the guidelines that are submitted.

So here in 1.1C.4, we're asking for information around the guideline and whether it does support the measure focus. Here, we'll ask for the specific guideline number or page number, a quotation of the specific guideline recommendation verbatim and the grading of the recommendation with the guideline. And here as an option number one, the guidelines actually provide a systematic review and grading of the body of evidence and this – this one option for developers to actually be able to submit evidence and here what they did was actually look at the guidelines and then actually provided the systematic review of the evidence.

There are two other options that the measure developer actually provides – does an evaluation – or looks at the literature to describe a systematic review of the evidence that's already in the literature and the third is that the measure developer actually does a systematic review of the literature if a systematic review is not currently available. We expect that when we're looking at findings from a systematic review that they're not simply just descriptions of individual studies but that they include quantitative data and a synthesis of the entire body of evidence.

And I think that kind of wraps up as we're looking at evidence – the three major ways that we would expect that measure developers actually provide evidence for evaluation by the steering committee. I will just stop there. I know there's a lot of information. Let me just take a quick scroll through here to see if there's any other areas that I should cover. But I will – I will just – let me just open it up for questions at that point.

I know that the evidence portion of the measure submission form and measure submission evaluation process often derives question. So maybe we'll – let's just open it up for any other questions that you may have related to the evidence.

OK. I'll actually just scroll back here. I just want to – in review – I really want to just make sure I'll go back here to page 25. You know, I went through many of those examples but I wanted to just bring this back up a few – at a higher level to just kind of recap what we've walked through.

Basically, what you'll be evaluating in the stage 1 is going to be importance to measure and report and that's going to be focused on three subcriteria, which will be high impact, performance gap, and the evidence to support the measure. High impact and performance gap, we'll use a generic rating scale of high, moderate, low, and insufficient. Evidence, we'll use a little bit more of a grading scheme in which you're looking at the quantity, quality, and consistency of the evidence and in order for the subcriteria 1C to (cast), quality, quantity – quantity, quality, and consistency of the evidence need to be rated moderate to high across all three.

If the quantity of the evidence is rated low but the quality and consistency of the evidence is rated moderate or high, this can pass the subcriteria 1C but only if it's judged that additional research is unlikely to change the conclusion, that the benefits to patients outweigh a harm, since there isn't sufficient quantity of data of evidence supporting the measure focus. If the quality of the evidence is rated low, it can pass only if it says that there is a clear potential that benefits patients over any harm.

If there is low consistency of the evidence, this subcriteria will not pass and there is an exception to this evidence rule, which is for health outcomes; however, we will not see any of those types of measures in this project. So in summary, I'll just point out that these are the three major subcriteria that will be evaluated in this stage one of this project, which will again be high impact, performance gap, and evidence.

And I'll just open it up to questions at this point and if Ashlie or (Lorelei) or Evan, you have anything else to add, feel free to just kind of jump in here or Andy, if you have anything else to add here.

Operator: Ladies and gentleman, if you would like to ask a question, press star one.

You have a question from the line of (Bob Ram).

Taroon Amin: Great.

(Bob Ram): I'm just saying (inaudible) I was trying to ask questions when you asked before but was not permitted.

Taroon Amin: Interesting. OK.

(Bob Ram): So I just want to let you know that.

Taroon Amin: OK. Thank you.

(Bob Ram): You know, a lighthearted one, if there's some pedigree of guideline that most people would agree, kind of pass the test for us not to have to replicate, describing what went into them, I'm thinking in particular of U.S. Preventative Services Task Force Guideline or Recommendations and I just

really would like to challenge the thinking where we have to go inside the study for let's say an A recommendation from the task force and repeat and kind of decipher on our own what the evidence-based review has already done.

I don't understand it. It just – it's concerning to us. We're faced with ever increasing burden and responsibilities around large measures that – so I'd like again to have response to that if I could.

Taroon Amin: I think the – I mean the – this was a request and guidance that we got from the Measure Evidence Task Force that this is the structure in which evidence should be submitted to the steering committee so that they have sufficient information to actually evaluate whether the measure focus is actually based on the evidence cited in the guideline. The reason why it's asked for and is fashioned in the measure submission form is so that the steering committee has sufficient information in order to do the evaluation.

You know, we don't expect that the measure developed – I mean that we don't expect that the steering committee is going to go into the guidelines and specifically do an analysis to ensure that the measure focus and the measure specifications are built on the specific methodology of the guidelines itself. So that's the – that's the perspective; however, clearly we recognize the increasing burden that we continue to raise the bar on for developers so that burden is acknowledged.

(Bob Ram): Why – I'm not sure I got a response to the question. The question was, you know, let's take a classic task force guideline like cervical cancer screening, ages X to X, possibly maybe with a – you know, you can skip and if you have an extra screening for X number of years. Just keep it simple.

You know, I don't know if that requires a lot and if it's an (inaudible) recommendation. I'm not sure that requires a lot of translation by the steering committee. I think they understand that. It's pretty straightforward. I guess I'm asking in specific cases and you considered doing some deeming around particular guideline if they base it on the quality that they do. I can understand the guideline doesn't equal a guideline, doesn't equal a guideline

but I'm really hard pressed where in fact these recommendations are actually entered into law.

So it's currently perplexing. I'm just thinking that if you are looking and addressing – trying to address the burden question and also address kind of like some fairly straight forward. There are occasions from fairly straightforward kind of guidelines to measure crosswalks. That you might want to value yourself and think again about whether or not that strategy is purposeful or whether or not there could be some – a simplification and make it in those condition – this work had been easier.

Taroon Amin: Well, again, we appreciate the feedback. I think the reality of it is that we are responding to specific guidance from the evidence task force and specifically noted that as you describe, the guideline is not – a guideline is not a guideline and guidelines are not sufficient for demonstrating evidence. We still expect that there would be an evidence review in the guideline and that would be something that we expect to delve or to demonstrate – simply citing guidelines is not sufficient in our criteria as it currently stands. And we're happy to take this conversation offline and have some, you know, more deep discussion with other but that's where we stand right now and that's where we stand in the criteria at this point.

But if there are any other questions from the steering committee members, specifically related to the guidelines – I mean to the criteria – I'm happy to take them.

Any – I know as the chair of this committee, if there's anything else that you kind of wanted to add or have me clarify, I'm happy to take those clarification question.

Male: Yes, can you hear me?

Taroon Amin: Yes, I can.

Male: I didn't know I had to star one.

Now, I think the criteria here are straightforward. I certainly understand Bob's frustration. I've had the same frustration myself but I might just say very briefly that there are probably only a few circumstances where guideline setting organizations are confirmed evidence based and evidence criteria – similar to the USPSTF that even would be considered to be deemable – if NQF ever decided to do that and I think that's probably, you know, it's very hard for them to draw that line. So I certainly understand your point, Bob and I kind of see it and – but I'm not sure that it's – there's a practicality of which guideline setting organizations are evidence based enough. So that's a very tricky thing. But I understand the stance that's taken today and that evidence-based body could potentially debated against some time.

Taroon Amin: Are there any other questions from the steering committee members?

All right. Well, I'll just wrap up by saying again, you know, my name is Taroon Amin and we're supported by a great group here with Ashlie Wilbon and (Lorelei) and Evan so if there's any questions that you have as you're beginning to receive concepts for review about the, you know, related to the criteria, feel free to please reach out to us. We're here to assist you in any way possible as you start to evaluate measures.

And we're very much looking forward to meeting all of you here in person in Washington next month. So again, thank you for all your time and we look forward to meeting you soon.

Male: Thank you.

Taroon Amin: Take care.

Female: OK.

Male: Thank you.

Male: Thank you.

Operator: Ladies and gentlemen, this does conclude today's conference call. You may now disconnect.

END