National Quality Forum

Moderator: Sheila Crawford
January 29, 2014
5:00 p.m. ET

Lauralei Dorian: Good afternoon, everyone. This is Lauralei Dorian from NQF. First of all, please forgive my voice. I have a bit of a cough and a cold. I just wanted to check to see who we have on the call today.

(Jennifer Layo): This is (Jennifer Layo).

Laralei Dorian: Hi, (Jennifer).

(Jennifer Layo): Hello.

Laralei Dorian: I think we have -- is (Barb Gage) on? I believe (Barb) and (Juan) and (Charlie Lincoln) are streaming the webinar. So they'll be able to hear it. It's just 2:01 now. I'm going to hand it over to (Angela). This is a very informal call, so she'll talk to you.

(Angela): So first of all, I wanted to be sure can everyone on the call hear me all right?

(Jennifer Layo): Yes.

(Angela): Great. Perfect. And if there are folks who are streaming the webinar and want to ask a question, you will have to actually dial in to ask the question or type in your chat box to ask a question. OK.

So the purpose of today's call was really to check and see as steering committee members are starting to think about the work, whether there were

any questions after reviewing your committee guide books and supplemental materials about measure evaluation.

The second thing we wanted to do was quickly walk through a measure and help the committee with thinking about how they should be evaluating measures as they go through this process. And a third thing was any questions about our process and I think we'll conclude with our next steps. So that's our agenda for today.

So starting out, does anyone have questions after they've read the materials thus far? OK. Well, then hearing none, I guess we'll dive into an overview of one of the measures that's been submitted to our projects.

I will preface this by saying that staff is -- staff are reviewing all the measures that have been submitted to our project and will be providing a summary of the measures to assist the committee in their evaluation of the measures … calling out questions, letting you know what was submitted, what was not submitted if that's the case, and helping tee up the measures for you as we walk through the process. And those measures will be available for you to start reviewing in the next couple of weeks.

So let's move on to the measure that we have for today, which is measure number 487 EHR with electronic data interchange describing use in encounters where a prescribing event occurred.

You should be able to see that up on your screen now. OK. There it is. OK. So I'm going to spend as I walk through most of our time on the evidence and measure testing as those are the most complicated parts of the measure.

So walking through the form, the first piece of form you'll see is the brief information and in our staff review of the measure, there's also some other information that you should keep in mind as we walk through.

It is -- it is a structure measure that is it's related to using electronic data exchange and the rationale provided in 1B1 of the measure form is that

measuring this structure will have the impact on outcome, which is fewer errors and dispensing.

And please note, that the majority of measures in our projects are structured measures for this current round. Also, please note that this is a previously endorsed NQF measure or maintenance measure.

It's also a measure that's in use in the Federal Meaningful Use Program. That information is later in the form, but it's something to keep in mind as we walk through.

So moving on, further down the form we're looking at, again, the measure type of the structure, the data sources that were going to be used, electronic clinical data, and currently this measure is specified for at the level of analysis of individual clinicians, and that's going to come into play a little bit later as we walk through the measure.

So other parts related to -- you can see in B4 whether the measure is paired or a composite is not applicable to this measure. Next, we move to the evidence portion of the measure information form and for 1A, this section is actually an attachment that you'll have to pull up in order to see and evaluate what the developer has provided. So we'll pull that up for you and walk through that.

OK. That's a -- OK. So what you'll see is the top portion, which should be completed with the measure number and title, that's missing here, but that's not a problem for the substance of our evaluation.

You can also see NQF has laid out specific instructions and a special note is the second box just down here in the note box. And it provides the developer and committee specific guidance relating to how to evaluate the evidence presented to support the focus of this measure.

So for this measure, which is a structure measure, committees should be looking -- committee members should be looking at whether the developer presented a systematic assessment and grading of the quantity, quality and

consistency of the bodies of evidence that this particular measure leads to the desired outcome, which is fewer errors and prescribing.

So moving down -- just squeezing down a little bit more to 1A1 -- a little bigger so it's easier. So 1A1 (inaudible) measure. Just continue to scroll down to 103. OK. We're asking the developer to present information about the rationale for the measure, and specifically we ask them 1A3 the source of their systematic review of body of evidence.

They selected other. Generally, we like to see other systems for grading evidence as laid out earlier, but they have other checked. So we'll skip down to 1A. OK. Other source of evidence.

All that is provided here that the measure's in use in the Federal Meaningful Use Program and what we really would like to see, and what we typically ask developers, if they were to lay out, as I said before, the quality, quantity and consistency of the evidence, what we want to see here is a robust description of the evidence on which the developer is basing their performance measures.

That doesn't seem to be present here and we do provide examples of what good looks like regarding evidence on our website and we'll also be providing that to you if you get your package of measures.

So taking what you have before you here so far, we want to review our first algorithm related to evaluating evidence and follow the pathways as described on our orientation call and that will lead you to a rating based on what's presented so far on this measure.

But keep in mind even as you're going through, committee judgment and expertise should also be applied to determining how you rate the evidence.

So some questions after going through this portion for the steering committee might be whether the evidence is directly applicable to the process of a structure being measured and whether this structural process is proximal to outcomes still intact.

And you can also determine whether an evidence is (inaudible) to be applied and go through the analysis again as described in the algorithm of whether there could be performance measures that are based on better construction to lead to the desired impact and hear -- the second question would be around whether a systematic review or assessment is presented and here, it's not.

So that will come out of this attachment then … it's one of the shorter ones, unfortunately. Pulled a measure at random, but we'll come out of this attachment and go back to performance gap on the main form.

So here, performance gaps. Here, we want to see, as we say, the demonstration (inaudible) performance quality problems, so the opportunity for improvement; and a brief explanation of the rationale that illustrates this gap in performance of suboptimal performance on this measure.

Here, the provider or the developers provided a statement that more electronics prescribed and leads to more legible prescriptions transmitted securely leading to less errors and dispensing. And no other information is provided in one or two or three or four. And this is a case where we would, again, ask the steering committee to use their expert assessments and experience.

So let us know if there is additional information that could be provided out there in the field about performance gaps and specifically whether there's disparities data available or performance data available that's relevant to the measure focus.

So after reviewing 1B gaps and questions the committee might ask of themselves would be … should this measure be satisfied with disparity? So we'd ask that you let us know that as well.

So moving down to our high priority criteria, which is previously referred to as high impact and this is whether a specific national goal is being addressed by this measure. I'm looking at this and (inaudible) describing errors of a large problem in 1C3 and in 1C4. They do provide a citation for data demonstrating that this is a high priority area.

And staff as we review will be pulling this citation and one of the things we'll be looking at is whether the citation does support the statement in 1TC. This is a problem prescribing areas that can be addressed by electronic prescribing and present that to the committee.

1C5 is not applicable. So just to recap on 1C -- on 1C, here we were looking for a brief summary of the impact data and the question the committee might ask themselves would be does the measure address the significance of problems and whether the data presented or citation that's presented will speak to what the measure is trying to accomplish and support what the measure is trying to accomplish.

Laralei Dorian: (Angela), we do have one question from (Charlie Wilson)…

(Angela): Yes, great.

Laralei Dorian: … wondering whether the NT staff ever go back to the measured developers to ask for additional information or more detail?

(Angela): Yes, (Charlie), that's an excellent question. You know, we're starting with a new process in this round of measure evaluations. In the past, we definitely would go back to the measure developer and ask for additional information.

In our current process, we haven't approved detoured timelines. We haven't the time. Nor is it -- it's something -- nor is it something that we're encouraging based on fairness to other developers, but it is something that could happen if the steering committee specifically requests additional information I believe from the -- from the developer.

But we have done that in the past, gone back and asked for more information be submitted, and that's a possibility here.

Laralei Dorian: One thing maybe when you come together on your work group calls, and dive deeper into these measures, if there's something that you think we really need in order to recommend or not recommend the measure, there is the opportunity for sort of (fly back) to the developer if they can come back in that time sort of input.

(Angela): I'd also say that to hear this criterion, number one, evidence is a must pass criterion, and if the steering committee feels it's warranted pulling back at this point and asking for more information, certainly a possibility.

So moving on to the -- to our form, we're looking next at our next second must pass criterion, which is scientific accessibility that's based on the reliability and validity testing. And we're looking at the attachment here for 2A2.

So this is another point at which you'll have to reference the second document, the measured testing document. Again, this is a structured measured and we've laid out instructions specifically asking for a brief summary of the specifications.

We want to be sure that the beta elements are clearly defined, codes are clearly listed and enumerated on this attachment including any definitions that are going to be needed to interpret the measure, and specifically when I mention codes, (inaudible) 10 codes should be included if they need to implement the measure.

Then -- and also we ask that developers indicate whether the outcome of this has risk adjustments that (inaudible) measure.

So scrolling down here, we've noted what it means and what we mean by testing and let's go on down to -- it's past the notes section to 2A2. Liability. Yes, reliability section.

So here the developer has indicated that they're testing the measure at the performance for -- measures for which would be a signal to noise analysis, which would be provided. And we note that for 2A2 how the measure is then calculated, how the testing has been done, which is the average electronic prescribing rate calculated each month for the same practices.

Let's see here. And I apologize. Actually, before we got to 2A2, it should be noted that this data was gathered in 2010 and that's in Section 1.3 and the data was detected and scattered over 6 months in 2010.

Also, we would've noted that testing occurred at the individual and group practices level, which matches the specifications in part referring back to the level of analysis provided in this measure by the developer where in the level of analysis, they said they suspect this measure is for individual clinicians.

So the measure is currently technically on what the level of analysis is, but especially in submitting my asset developer is whether they intend to specify the measure both for clinicians, individual and group practices. So that's something that might go back to the developer for clarification since they have the testing to support that.

So again, focusing on 2A2, for each level checked above describes the method of the liability section and as we've said, the developer here provides a description of how they come back with the testing and 2A23, they also describe the results and they also include 2A24. Their results -- interpretation of their results.

And here, we might have a question and we have a methodologist look at it. They state with the large sample size, we had results indicating that the increase observed per month was real consistent and not due to noise.

Typically, we like to see numbers here and then descriptors about the noise -- signal to noise analysis and looking at this right away, we're not certain if they've met that requirement. Again, as I said earlier, we -- staff will be reviewing these and providing our thumbnail analysis of whether or not this meets our requirements.

We'll indicate whether the type of testing is appropriate type of testing and provide that information.

So just following after looking at 2A2, the questions for the committee here might be are the data elements clearly defined, which I'm not sure that we answered that question looking through the form. We also want to know whether all the appropriate codes are included. There are no codes included for this measure, but that would be a question for this developer.

We'd also want to know is it likely that this measure can be consistently implemented? Keeping in mind, if this was a measure that is in new -- in the Meaningful Use Program, it might be that the developer simply needs to provide us more information about these questions.

And then you would turn to algorithm number 2 rating the measure testing in terms of reliability and follow that pathway to determine how you can rate the measure testing presented by their developer here. Again, this one is a little more objective, but its objective judgment is also encouraged, too.

So moving on onto the same form to validity testing, this told us it is a performance -- validity testing was performed -- actually performance measures for and it's for the systematic assessment of state validity.

And they here starting at 2B2.2, they described the method of validity testing and the description here reads, "There were no comparable rates to compare results to. The average rates collected matched broad numbers found in the literature. Spot check was done at a few practices through an electronic chart review to validate results of the numbers determined in the chart review appears to match automatic aggregated number (inaudible).

This … without, you know, kind of weighing in, this is exactly what we would be looking for in terms of systematic assessment of space validity. Again, we'll see running this through our staff review and providing additional input about whether this meets validity testing requirements.

Moving on to what were the statistical results from validity testing. We like to see numbers here. What is provided here is that they did calculate averages on practices that were next -- verified with the manual chart review to cancel cases, but for our purposes, we'd want to actually see those coupled averages included in the form.

So this is maybe another case where we would go back to the developer and ask for that information.

So moving on, is -- are there any questions?

Female: Yes, I have a question.

(Angela): Yes.

Female: Please. The criterion on the form emphasizes space validity. Is our review limited to space validity or are we also looking for other types? Because space is pretty low.

(Angela): Space? Yes, we consider it low, but this is all that the developer has provided. We allow for various types and encourage various types of validity testing, but this is …

Female: But in our review on the materials that we reviewed earlier this week...

(Angela): Yes.

Female: ... it being the algorithm is space validity, so in our review, is space validity sufficient for a, you know, kind of a pass at this stage or would we be looking for higher levels of validity testing?

(Angela): So I think space validity -- and I'm looking at the algorithm, which we can flash up, it's -- looking at algorithm number 3...

Laralei Dorian: Page 13.

(Angela): ... Page 13 and I -- the space validity criterion can be -- can be -- can lead you down the pathway to radiant for a high if it's conducted with a computed performance measures for, but here, we really are looking at not a computed performance measure score.

So if you're looking at your algorithm 3, box 3, leads you -- this is space validity. So then you go straight to space -- I'm sorry -- box 4 and the highest rating coming out of that would be moderate, lower and sufficient. Sorry. One hundred or lower is insufficient.

And you'll see in box 4 we have a specific -- a specific method that we're seeking in terms of rating assessing space validity and that does not appear to be applied by the developer here. Does that answer your question?

Female: Yes. So in general, if we follow the algorithm, we should be fine?

(Angela): That's correct. That is correct. So all right. Where were we? We had moved on to 2B3 and 2B24, your interpretation of the results. Again, we would want to see numbers here to support the statements.

So just again, following your algorithm may not lead to a rating of high for this particular submission as it stands right now. So 2B42, there's no exclusions for this measure and no risk adjustment of stratification.

So moving to 2B42, they have a statement about how they pull the data. We used all available data from practices and not a sample. There's no recent stratifier excluded in these cases. Measures reflective of the activities of all providers, which is what we measured.

So reaching the end of this particular section, the question from the steering committee might be whether given the data and information that the developer's provided, this measure is -- does this measure meaningfully -- does it identify meaningful differences in quality among providers and that is the area where the committee has to apply both of the algorithms as we discussed earlier as well as their judgment.

So scrolling down a bit further, we'll look how missing data was addressed. And they tell us that they were able to determine the zero result was from no electronic prescribing or zero was from missing data by checking the transmission.

And a few practices were a true zero as was this was verified by calling the practice and talking to the vendors, but most cases of zeroes were due to missing data. And the way they handled the frequency of missing data was 20 to 30 percent and that's a number that is -- sorry -- a committee might question the impact of that missing 20 to 30 percent of data due to transmission issues at the site.

And then they talk about how they interpret the results to show that the performance results are not biased and they state that statistical analysis

weren't done to confirm how the biases were handled. So that's, again, when we turn to algorithm number 3, a pathway that may not lead to a rating of high, but potentially moderate, low or insufficient.

So moving back to the main form. We're looking now at feasibility, which is down -- number 3, feasibility, our third criterion. This is particularly -- this is the criterion where, again, we rely on the steering committee to discuss the issues around feasibility of capturing or collecting data required for implementing the measure and whether there's a burden associated with it and whether that burden is an undue burden.

So here is 3A, data elements generated as a byproduct of care processes, and it looks as if data for this measure are generated in the normal course of care. So collecting the blood pressure lab value is another, which would indicate that there's a very low burden there.

Moving on to 3B, the data elements are especially specified in electronic health record defined fields and this is -- this is not an e-Measure. So this is not data. There is no feasibility assessment that is (inaudible).

For the data collection strategy, we don't have … we should have a discussion in here and they have left it blank. So that's a question for the developer as this measure comes under review.

Fees and licensee issues do not -- do not apply to this measure, so perfectly (inaudible). Moving on to usability and use. They've laid out for us the current use. This is the maintenance measure. Again, this is a maintenance measure that's being publicly recorded, which would indicate that it's both -- it's usable.

And we also wanted to note here if there's been progress in looking for the -- if there's been progress on the measure since it's been in use and looking at what's been provided, it looks as if the e-Prescribing rate has been improving 10 percent year-over-year and the average rate of performance is at 80 percent.

So this indicates very good progress; however, a question that the steering committee might want to ask here is whether this will be a measure that is

going to be soon topped out or at the top of performance and that might be also a question for developers.

Keeping in -- so keep in mind that we did identify two issues going through the two must pass criteria, which are the evidence provided and the testing provided, and typically if we were evaluating this for real, we wouldn't -- might not be able to get past those two and feasibility and usability would not be looked at.

So really, the emphasis should be on making sure those parts must pass criterion are really met by both measures.

So assuming the measure made it through all of these criterion and were found to be acceptable in terms of ratings, we would then go and look at a comparison of other measures that are in the portfolio and determine whether they're related measures.

So whether that they meet our competing measures, and so whether harmonization of the measures, the measure before us in this project and measure in the portfolio would need to be harmonized.

Do you want to talk about that one or ...

Laralei Dorian: Sure. I'll try to talk if my voice doesn't give out ...

(Angela): OK.

Laralei Dorian: ... because of this cough. Yes. So I don't know how familiar you are with the idea of related and competing measures. It was started a few years ago because we certainly don't want duplicative measures. We don't want more measures that are necessary out there in the field because we bring that system to people.

For some measures, it's important that there are two different measures, even if they're very similar because they might be specified for a different setting, one involves children and one involve adults and there are certain clinical

reasons for that, but there are other measures that actually are looking at almost exactly the same thing and we call them competing measures.

We actually, what we tried to do at the beginning of the project is go through all the measures to identify the database competing measures. We don't have any competing measures in this current project. We do have a number of (inaudible).

So an example of that is one that's called adoption of medication we're prescribing, which is quite similar to the one (Angela) just walked you through. So, as (Angela) stated, what will happen is that we will review every single measure individually and only then if they're recommended will you talk about the database competing.

So in this instance because they are related, we asked the developers to, what we call, harmonize the measures, so that they're as comparable as possible so that the results that you're getting can actually compare them as (inaudible) measures because something like that might be … if one measure is specified for 15 and up and one is specified for ages 18 and up and we would ask the measure to (inaudible) ask you what it should be (inaudible) what would make more sense.

And then whatever you recommended you would go back to the developer and say please harmonize these measures, make sure you have the same (inaudible) or the same exclusions. So that's the process of related and competing measures.

We're aware that it can be confusing with some of the measures out there and we certainly don't want to endorse more measures than are necessary. And the ones that we do endorse if they are necessary for more than one, we want them to be at similar as possible other than their necessary differences (inaudible).

Are there any questions about related and competing?

Female: Lauralei, just a practical question. When we identify something as related or competing and the request goes back to the measure developer to work with the other developer to harmonize, what is -- what's the success rate of that?

Do the developers usually work together and resubmit?

Laralei Dorian: Well, that's an excellent question. Traditionally, it's been a very haphazard, confusing process because what used to happen is that we would go -- essentially harmonization would be discussed for the first time during the in-person meeting.

The developers had not had a chance to talk with one another and we didn't really have clear guidance on what would happen following the in-person meeting. So about a year ago, we rejigged, I guess, our harmonization process; you have a Lean event and one of the major items that came out of that was for SAP to reach out to developers much earlier on in the process. So essentially now for this project.

And the project that I was just on actually with (Angela) as well, behavioral health, was the pilot project through that and we had great success actually and we offered to facilitate calls and all of the developers under that saying, yes, this is great. Thank you for getting to us earlier. We're willing to make these changes and that's been replicated in other subsequent projects as well.

So we're hopeful that that major change that we're reaching out earlier and offering to facilitate calls and that that's leading to more collaboration between the developers.

Female: That's great. That's a wonderful process. And I wondered about it in previous reviews that has been an issue that it's a lot of work and I just wondered if there'd been more success. So that's really good to hear.

(Angela): Right.

Laralei Dorian: Thanks.

(Angela): So just to throw the floor open for questions, additional questions about this process -- no?

Laralei Dorian: I just -- just to note that when we do send you the measures, I think, you know, a lot of it -- the purpose today was to walk through a specific measure, but I think we need to sit down and spend some time with your first measure if you're not use to this -- not used to exiting out the process.

It will slowly start to make sense and fiscalize and you might come up with more questions at that point at which time we'd be happy to answer them via email or over the phone.

That's really what the work group calls are for as well for you will have reviewed a small subset of measures and on those calls, we'll be there to answer questions and to talk through, you know -- for example, some of you might have thought the evidence was high and others medium and we can talk through what that process is.

(Angela): That's exactly right. And also, when you do get the measures -- the actual measures before you and the measure summaries, we'll have teased out a few questions for you and would welcome any additional questions or dialogue certainly before the work group calls or you can reserve them for the work group calls.

But any time you have a question about the measures, let us know.

(Dawn): Is there -- this is (Dawn). Is there a specific person that you would want us to go to or just kind of call out the project lead?

(Angela): That would be Lauralei or myself.

(Dawn): OK.

Laralei Dorian: Yes, we're here...

(Angela): Yes.

Laralei Dorian: ... all the time.

(Dawn): Thank you.

(Angela): Are there any next steps that we need to...

Laralei Dorian: The next steps are the same as when we discussed them on Monday, which are the two upcoming work group calls. We will let you know those dates by the end of the week. You've already been given the dates. You should have three or four holds in your calendar, but as we discussed on Monday because we have fewer measures than a typical project has, we'll probably only need two of those calls.

So don't release the dates yet, but give us a day or two to determine which are the best dates, and then we'll break you into two groups and put the measures out. And as (Angela) mentioned before, we'll be sending all of the measure information out to you within the coming weeks.

And just to refresh your memory, those work group calls are at the end of February.

Male: Sounds good.

(Angela): Great.

Laralei Dorian: OK. Well, we're here to answer any other questions if any come up.

Female: Lauralei, I have a call on Friday; is that happening?

Laralei Dorian: Yes, that is happening.

Female: OK. That is happening then? All right.

Laralei Dorian: So that's actually just a replication of the call today. So there -- both of these calls were optional and you certainly don't have to attend both of them because we'll be going through the same exact information; although, there may be different questions. So...

Female: Thank you.

(Emilio): Hi. This is (Emilio). I just to comment that with the algorithm and these work flows, I think that it looks like -- process is a lot simpler and more straight forward than a couple of years ago. Good job. Thank you.

Laralei Dorian: Thank you.

(Angela): That's good feedback.

Laralei Dorian: Good to know. And I think, (Emilio), what you'll also notice -- (Angela) was talking about it -- that we have those staff reviews that we will have some staff questions and a few short comments about each measure. So that's different as well and we're hoping that'll be helpful guidance.

(Emilio): Very good. Thank you.

(Angela): OK.

Laralei Dorian: Great. Well, we can give back some time.

(Angela): Yes.

Laralei Dorian: We'll give you back some of your afternoon, which I'm sure you're happy about.

(Dawn): Thank you.

(Emilio): Bye-bye. Thanks.

Female: Thank you. Thank you very much.

END