

NATIONAL QUALITY FORUM

Moderator: Reva Winkler
March 11, 2013
11:00 a.m. ET

Female: (Dave), scroll down where the blue goes away and the – there's the non-colored sections that's where we're going to begin. OK, Mike?

Michael Lieberman: Yes, that sounds good. So, we're going to start on line 25 of the spreadsheet.

Female: Right.

Michael Lieberman: Yes. OK. Well, good. It looks like with the WebEx we have enough today, so we can kind of read over the recommendation and discuss it. So, we're on – this one from American Urological Association about the data feasibility.

Is everybody on the WebEx? Or should we go ahead to check – I can go ahead and read it anyway I guess. Data element feasibility assessment data in a structured format deserves further consideration. The current availability of structured data is largely dependent on the quality and prevalence of vendor-developed templates. Providers may find a suitable template that can capture information structured fields with one EHR product but may have to input this data by free-text in another due to lack of available structured fields.

Furthermore, existing templates that facilitate structured data capture may have poor usability; so providers may elect in certain circumstances any relevant data more quickly in narrative form. Therefore, feasibility assessment of data elements should not solely evaluate any eMeasure based

on the presence of existing structured data fields, but rather on the type of data needed and suitability for capture and structured form.

Paul Tang: I don't ...

Howard Bregman: This is Howard – go ahead, Paul.

Paul Tang: I also know that we – I think they're asking to allow like unstructured text as in the dictation to account and I don't think that would make it more – I think that would make less feasible, right? Did I – am I reading that correctly?

Michael Lieberman: I think that's part of it and they say, you know, but rather on the type of data needed in the suitability for captured in the structured form.

Paul Tang: I see. So, you're saying if you could is that – is it still efficient to do so?

Female: So, to me they're mixing up (inaudible) concept here. I mean, I think they have to – I think we have to do – have to make it clear that we want structured deals and if they – and again, they're mixing up from workflow issues to but the issue of basically transmuting text field into structured field, I think we need – to may have a little paragraph about and that that may be possible but we need to, you know, that needs to be worked on, you know, by the gurus, if you will.

I mean, one option to have a pick list and – plus the pretext but – I mean, that's, you know, there are many ways to skin that cat including natural language processing and things like that.

Paul Tang: I the prototype of this is this whole (smoking) status. It makes them so they'd be captured in structured form but, you know, in systems where you have to click five times, I wouldn't consider that feasible. I think that's a bit of what they're saying and it goes back to usability of the EHRs. We want to make sure that vendors understand that it's not just presence or absence of structured field if that has to be very accessible during the usual workflow.

Female: Again, I mean, I'm reading ahead to the next comment. That's a workflow issue.

Paul Tang: Yes.

Michael Lieberman: Yes. I think we've captured that issue as part of workflow. So, I kind of agree with the – and wait – and when they say – but rather on the type of data needed suitability for capture – yes, either – that's subjective request or I mean, I think it is capturing workflow.

Paul Tang: Another word, we're going to create (inaudible) comment in some set that we can get ...

Michael Lieberman: Yes.

Paul Tang: ... you know, that it has to be efficient and natural, not just coded.

Michael Lieberman: OK. Let's move on to the next one.

Howard Bregman: Can I make a few comments?

Michael Lieberman: Oh, yes, please do.

Howard Bregman: This is Howard. When I think generally when we – from the – well, I think the vendors and the providers say something is feasible, they just mean that – it doesn't mean that there is an existing form which captures that information. I think they mean that like you could have a form and if a user used it then it would be reasonable to capture in their workflow because they already captured that information.

So, we would not say, and I don't think a provider would necessarily say we already have a form for that, so therefore it's feasible. Anything could be captured in a structured way in usual clinical documentation that could be used for quality measurement most of which is not captured into in discrete form and could be – if they were to make everything into a form.

So, you know, I think there's an issue of when we say something is feasible, does that mean, "Oh, yes, right now, today, we already captured it in discrete form and no modification at all would be required."

Keri Christensen: This is Keri ...

Howard Bregman: I think that's probably too stringent of a definition.

Keri Christensen: This is Keri Christensen from the AMA-PCPI. And I agree with you Howard, but I add the caveat that on – we need to be very careful about what we're talking about the feasibility as and we're talking about the feasibility of the measurement.

So, to some extent to say something it's feasible today. to me, that means I can go calculate this measure today on as opposed to the feasibility of an EHR collecting information which is something slightly different.

Michael Lieberman: It seems that there is something is already captured in structured data today in a, you know, and meets the other criteria that we talked about and that it in a standardized workflow that doesn't put at undue burdens and whatnot – one can say that that's feasible and the data element would score well but that is not the only way that the data element can score well in this.

I mean, we have in here the issue of what's going to happen in the future. We have some – so I think this is already kind of have been addressed by the overall framework that we have and it's, again, you know, just being able to measure it now. So, it's a lot about the fact that it might be feasible but it doesn't mean that if you don't capture it now, it's not feasible.

Paul Tang: This is what we meant by the two columns, you know, now and then three or five years?

Michael Lieberman: Yes, I think that's part of it as well.

Howard Bregman: I will – let me give you an example and see what you think it is. I am – I could type something that has meaning and because I'm typing it it's not going to be captured in discrete way. If I stop typing it an instead I used a tool that I created which I would have to remember the use that would pre – would fill in my note with certain text and then that would be captured – if I use that tool, it would be captured in discrete way, but if I just did what I usually do which is type, it wouldn't. Would that be – if someone were evaluating that

situation would they give that a high score for feasibility or a low score for feasibility?

Paul Tang: Well, Howard ...

Michael Lieberman: I think you'd go through the whole – OK, you go through the whole form that we have and you'd evaluate on that. So, if we would say, “You know, a difficult workflow that you created to try to capture that in a structured format or it would score low on workflow. It might score – if it was just a format going to a coded data element that met all the other criteria could score high there and that, you know, under data accuracy, I guess that's the question of whether we thought, you know, that it may not score – may score medium there as well if it's not workflow still is going to be used very often.

Howard Bregman: Well, I'd be interested in hearing Shannon's perspective on that issue.

Shannon Sims: I mean, I might – you guys know I'm a heartless pragmatist. I think that, you know, the feasibility (helpful) about what we can do right now from us, measure developers, I've got window, I think we've discussed previously as if – if it's possible in the future, it really depends on if the body that's either imposing penalties or incentives choose (inaudible) – I can't tell – they say they can't do that.

But our goal is to provide them a reasonable framework that assesses (better) but the fact that you can capture. I agree with Howard. I can capture anything in a structured way. It's just that I have to browbeat my providers into doing it and that makes a feasibility a lot lower and certainly the reliability down the road a lot lower because it's not going to (viable). I'm really forcing to do something that doesn't exist in their current workflow. It takes years for that to become (inaudible). Is that ...

Michael Lieberman: Hey, Shannon, I think, doesn't our – doesn't our system account for that under data accuracy where it says, “You know, three, two or one – two is the information may not be from the most authoritative stores and/or has a moderate likelihood of being correct and when you say, easy software to vaccination but that might also be the – if it's a, you know, if it's a – we're expecting providers to collect it in a part of the EHR that they don't use very

often, I would perhaps that would be – that would score two.” Likewise, if they would – workflow that required four or five clicks where they would usually just type something into their note, I think they would score low on workflow.

So, I guess my point is this – that I think we try to address this very issue in our scoring system and so that – to that – to be flexible to say that, you know, you can have that data on it but it might not score that well and it might be an area that needs to, you know, takes a little bit more focus.

Howard Bregman: I’m sorry. It’s Howard, I am disconnecting myself then I get back on.

Michael Lieberman: OK.

Howard Bregman: I think that the reason I am dwelling on these issues I think it’s an important point. I think we, as a committee, should have an idea of a given scenario whether it should score well or whether it’s should score low. And I’m not sure that we’re clear on that.

My – I think that we were pushing during our in-person meeting and in subsequent calls that the threshold for feasibility would not be very high. Meaning that it shouldn’t be that many, many things are scored infeasible and only a few things are scored as feasible. It should rather be a lower threshold. So, in fact, many things are scored – many existing things are scored feasible, but I think if we say that if I had to build a tool to capture discrete data because almost all my providers are just (writing) tax and then they would have to remember to use that tool.

If we’re saying that that’s an example of infeasibility, then many, many things are going to be scored as infeasible and we should – we should have a position on that.

Shannon Sims: Howard, I think you might have been disconnected but I thought Mike did an excellent job at filling how if you consider all of the parameters or attributes that – ones it does given indication of what is likely to be feasible now, and (can take) anyone in isolation.

The other thing is we do have to make progress. In your example, you could either dictate vital signs or put vital signs in or the (word) just decided that vital signs are very important and think most physicians believe that and we have workflows which stiffness (very) involved with physicians to make sure it gets structured on – in structured form.

So, we're making these kinds of judgment as we go through each of these criteria. They just can't be considered in isolation but I mean, I think we've gotten through this point many, many times and I really like the way it might get presented the way that the total scoring does account for this different considerations.

Howard Bregman: And I think what makes that is – that would – that situation would score low overall.

Michael Lieberman: Well, it would – it would score, I mean, it may be medium. It kind of depends on the specific data element we're talking about and it wouldn't score perfect that's for sure and nor should. But it doesn't mean that imperfect score, you know, it doesn't – an imperfect score doesn't mean there's something not feasible and we kind of have gotten away from saying either, you know, plus, you know, feasible or not feasible. It's a degree of feasibility.

So, you will have an opportunity to score an element that scores really high would probably be one that's like you said a vital sign or a lab result that we feel is, you know, it should score very high versus something that does require a convoluted workflow (inaudible) aren't going to use that (optimal) very low, but there's going to be lot things in the middle.

And actually, Paul's example of smoking status, how do you know, in my practice, smoking status is now captured as it needs to be for meaningful use and it's accurate. I mean, when I go through MAs, my medical assistants capture before I get in the room and click on the right button and then whenever – when I ask the patient, it's right.

I mean, so that's an example where it is, you know, there's a little bit more workflow involved in it in a lab results but it's not a highly accurate data on it that does include clicking buttons (inaudible).

Howard Bregman: If most of what were worried about were vital signs, lab results and smoking status, this committee would not even need to exist. What really – what we're dealing with and what all the commenters are dealing with are all of the other kinds of data that are stored in different ways and are more – much more complex in general than those elements which are very simple.

Helen Burstin: And this is Helen. Just to build on something Howard was saying (inaudible) (run) outside if it's (loud).

But I also wonder a part of what Howard is saying is whether we also need to scale the potential for an element to become feasible in some way that actually had some grounding in it and just as (the button).

Paul Tang: Well, again, I guess what I think we have two columns, one is for current, one is for near, you know, medium (inaudible) three to five year.

Howard Bregman: Well, I guess you hear is current and what we're really talking about in the future.

Martha Radford: This is Martha. I think we – also we need to throw a little bit of this back on the vendors and there may be – if there's any vendors on, please say something. I mean, there – we need to be clear that items need to be captured in some sort of structured format and that there's always attention between text and structured format and there may be innovative ways to convert text into structured format.

Some of that involves more workflow by the clinicians because they have to do two things basically pick a category and right text or you could have some of other message for making that easier for the clinician that almost certainly will involve natural language processing of some sort.

Sarah Corley: All right. Well, I'm a vendor on the call. This is Sarah Corley, and I'll tell you that I don't think that right now (NLP) is at a level where we want to accept the liability for it incorrectly populating something because, you know, the patient safety implications are huge.

Martha Radford: I would agree with that. I mean, I think that's fine but, again, this gets into now versus future and maybe it's not feasible now but we need to kind of work on (NLP) in order to make it more feasible in the future. And I'm just saying ...

Sarah Corley: But, you know, but I mean as a vendor our goal to put in our products, what our clients needs to get their work done in a (safe) sufficient fashion, meeting the regulatory demands but, you know, when we look at the tension between all of this, you know, reporting requirements is that our clients do not want to enter information that's not necessary for the provision of the case that they are giving at that point.

Now, we understand the need particularly in the primary care arena to take the opportunity when the patient is in the office to make sure that you were doing the services that that patient needs whether they're in for sprain ankle and they're not in for their diabetic follow-up, but we get a tremendous amount of pushback for that, and it's certainly easier for us to put things in a structured format but then you, you know, the whole note (bloat) and notes that all sound alike get pushed back as well.

And so I think that we – the whole health I.T. organization needs to really try and think what we're going to do capture the data elements that, you know, “Yes, so I'm sure with some research you might want to see the elements because it would be great for your project but we can't be crippling doctors that are struggling, we really need to, you know, right now do what part of the workflow, understand the feasibility what is commonly part of the workflow and what should recently be the expectation.

And the goal for the future certainly would be as we transform our healthcare system and allow for more team base care without fear of, you know, regulatory oversight that, “Oh, the doctor didn't do it himself; they can't get paid to collect more data elements than to have more patient collected data elements and certainly to mind using (NLP) but, you know, right now, it's a different world than what we hope will be the future.

Michael Lieberman: Hey, Howard, do you have a recommendation – I mean, we have a four-part strategy or evaluation system now with data availability (data) – the data standards and workflow to try to kind of assess whether or not, you know, the level of feasibility of anyone's particular data element.

And are you concerned that most of them are going to explore kind of the same low level and it's not going to be able – we're not going to be able to differentiate or is there an additional access that we should be looking at?

Howard Bregman: I think my main concern is that supposed somebody got this report when it was published, and they said, OK, now we're going to – now we have to do a feasibility – measure feasibility ratings. "Hey committee, you did this report." Here are five scenarios that we have to evaluate and we're going explain them to you in great detail, and we want – want to know from you in these five examples what feasibility score we should be giving.

I think that we should have an answer for that. Based on our discussions, we should have a general understanding about what this means so that if we were to look at these scenarios, we would all agree within reason what the score would be, and I'm concerned that instead what was going to happen (these were) – we wouldn't be able to answer that question in a consistent way because we're not – I don't think we're quite clear about what kind of feasibility scores ought to be given out in usual situations and these situations that anyone might be dealing with.

Michael Lieberman: OK, I think that's a fair – I mean, a fair request would be to – but that's probably, you know, kind of the – I think our process is then to get a system out there and to start putting it through – putting it through use and seeing how it works. So, you know, you have to start somewhere and I think that's what we've tried to do here now, getting – if it's five scenarios, you know, may be enough where I don't know if this group or another group or somebody that comes back together and looks (part) of this that, you know, and this is an example, this is how we would score it. Yes, I think that's fair.

Howard Bregman: Well, I'm not saying we need to publish five examples. Ideally, that would be nice, but we probably don't have the ability to do that.

Male: Yes.

Howard Bregman: But I think we should at least agree among ourselves what a feasibility score would be in different situations but I am concerned that the real world application of the scale maybe that everybody or that many, many elements get scored low which I don't think any of us think is a good thing. I think that we've actively discussed the fact that that would be a bad outcome if the end result would be that almost everything except the absolute obvious things like what we've mentioned vital signs, lab results, smoking status.

They're the only things get scored five and everything else gets lower score that wouldn't help with anybody.

Reva Winkler: And again, I'm just going to push back a little bit on – I agree with you. I completely agree with you. We need to have things that score reasonably high perhaps not in the immediate future but in the near future and, again, I would just say that I do find this vendor attitude around liability somewhat annoying and to sort of (take) the case around what is documentation and I think we all need to grapple with that and the issue that – I mean again, you have a choice.

You have an easy way to enter the data that you want to enter with text but then gets transmogrified into a structure then again there's many liability resistant ways that can be done including saying, "Hey, provider after you've entered here note, this is what's going to be entered in the quality of database." Is that what you mean and have them sign off on it? So, then you're – the vendor is out of liability issue. So, they sign off on it.

So or insert workflow, that's fine, we can say. Some of this is involved inserting new workflow and if it's important enough to measure quality, then maybe it's important enough to insert workflow.

Zahid Butt: Hey, Mike, this is Zahid.

Michael Lieberman: Yes. Yes, go ahead.

Zahid Butt: So, I think that it seems to me if I read the concern that they are expressing and, you know, I agree with, you know, what Howard was saying in the

context of, you know, the tool and things like that. But if we sort of maybe another way to do this is to look at their comment in a more sort of narrow sense, and it appears that they're may be more concerned about our statement that under the data availability which is within the sort of scorecard and also in the 2.1. Is the data readily available in structured format i.e., resides and fixed field of an EHR.

So, in other words, we sort of define availability only in structure than fixed fields and my guess is that that's what they're reacting to that it could be conceivable that in future tool or whatever else might come along and perhaps (NLP) might be advanced in us that someone could collect some thing not necessarily in fixed structured fields but perhaps in the codified format in which the value set as needed that in this case, it might imply that would be sort of infeasible in that state. So, perhaps if we could say simply that, is that data available or do we need to fix it that it needs structures quote/unquote, "fixed field" for it to be available?

Michael Lieberman: We could – I mean, I wouldn't mind getting rid of the reside and fixed fields with any EHR as part of the data availability but, you know, we'll just read it, the data readily available in structured format because I think, you know, as point out, I don't think anybody expects that currently, (NLP) is going to be able to take this the data for in and I mean this can be revised in the future. This doesn't have to be forever.

Zahid Butt: Or – and I guess if the feasibility assessment, it would say, "OK, this thing doesn't reside in any structure field and it's not available to any other means, so it's infeasible."

Michael Lieberman: We would score low in that part, yes.

Zahid Butt: How you would define it at that point?

Michael Lieberman: Yes, you know, we actually need to – we do need to move on, but I think that Howard from what I took away – I think making a recommendation that we – that we do come up with either more examples or I think it's a good one. Again, I don't know that it's within the – within the timing necessary to come out with this report, but that could be recommendation.

Howard Bregman: All right. I think you can move on ...

Keri Christensen: This Keri ...

Howard Bregman: ... and – go ahead.

Keri Christensen: This is Keri. I would be happy to take that offline with a small group of people. We've got some measures from one of our contract projects but I would love the (DataSwift) then if anybody wants to participate I'd be happy to start meeting them and we put some time together.

Michael Lieberman: OK. OK. So then – now, we're moving onto workflow so we've done one comment in a half hour. But so workflow per se, so we're on line 26 now, workflow per se is not characteristic of data elements, but rather may affect the quality of the data collected for specific data element.

In addition it will vary by EHR product and local implementations of those products, and even varying use patterns by individual clinicians. Workflow should be removed as a separate characteristic for testing, but instead should be understood as a mechanism that influences data accuracy and completeness, as well as location in the EHR and data format. The scorecard forces the reporter to assume that all EHRs and clinicians operate similarly; this certainly will not be the case.

Do you want to talk about that before the next? Or (inaudible) we'll look at the next paragraph.

Well, the overall data elements feasibility of scorecard document appropriately discusses two separate challenges to data elements feasibility. One is whether the structure of the EHR can support the documentation in the structured data and the second is whether the clinicians will document relevant data workflow. These are two very different issues with very different implications. I believe the guidance language provided in the Data Availability question in the scorecard combines the two concepts and is therefore going to provide unreliable results.

For example, for score of 3, the guidance states, “Data element is routinely collected as part of care processes and exists in the majority of EHRs.” For measures not achieving a score of 3, it will not be determinable if the low score is due to structural issues with EHR or workflow issues such that clinicians do not document in the structured data fields even though the capability exists.

Since the latter concern is addressed by the Workflow question, I suggest my – that the Data – I would say even longer, OK. The Data Availability ratings be revised to clearly represent only the issue of the EHR’s structural capability to capture the required data elements.

Male: So, Mike?

Michael Lieberman: Yes.

Male: I wonder if it’s (fast) to press the – either read off – going to read any print-out, I mean, just to save review time.

Michael Lieberman: Yes.

Female: Yes.

Male: And sometimes, it’s actually hard if we were (inaudible) (this thing) to you.

Michael Lieberman: OK. Fair enough. So, if everyone (to) finish reading that comment ...

Zahid Butt: So, Mike this is Zahid, while everybody is reading that, I think that this is similar to something we discuss last time. Wasn’t it? Where we discuss about the data element itself being sort of within workflow and then workflow as a separate and that potentially double counting the weight of it inadvertently...

Michael Lieberman: Yes.

Zahid Butt: ... I thought we decided to keep it the way it is. I think these are sort of addressing that simpler issue.

Shannon Sims: This is Shannon. I've chimed into that, you know, I think of what we're trying to do here is to build something analogous to the cage of the PHQ-2, we're trying to involve screening instrument but helps people identify potential problems early in the process and we've see further evaluation. So, my hope is certainly that the existing processes such as the committees that are formed to build these measures in an NQF endorsement process and (for the) comment would allow a deeper dive into the reasons that they might not be feasible.

So, I don't – I feel like we have processes in place to do a deeper dive. What we're trying to do here is build a screening mechanism that I think is pretty sound as of constructed and this comment doesn't dissuade from that.

Michael Lieberman: I would agree with Shannon. Anybody else have – want to comment on the other way before we move on?

Female: I would agree with that.

Michael Lieberman: OK.

Male: And I agree, too. Would that touch on the exception and I think we are trying to (fill) philosophy where want to minimize the exceptions because once you have to read the chart, you have to read the chart – the entire chart, so that ...

Michael Lieberman: OK. So, basically this is – doesn't have to be perfect but it has to be the tool to kind of uncover some issues with use feasibility for specific data elements and we think that we'll do that.

Female: Yes.

Male: I think ...

Michael Lieberman: OK.

Male: ... (inaudible) analogy to the screens were pretty (useful).

Michael Lieberman: OK. OK. The next comment is line 27, about the data of – go ahead and read that.

Any comments on that? I think we had addressed – we actually kind of addressed that under data accuracy where we say, you know, the likelihood of being correct which again, is somewhat subjective but if we, you know, if we feel that it's a field that's often cut and paste they did not reliably record it. They would score a little bit lower there as well.

Rute Martins: This is Rute. I think it goes back to Shannon's prior comment. This is a screening tool and so the copy-paste issue, it seems that the concern is more about data quality and that's not although it's a factor, it's not the focus or the single focus of this assessment.

(Christopher Millet): I didn't – (Chris) from NQF. I completely agree that this is, you know, with the screening mechanism comment and that this is good for kind of screening on accuracy but is there a reason why, and I know there's some other comments to kind of push on this but if there are reason why accuracy a part of what we're screening for, what we're checking for feasibility as opposed to when we're checking for (relative) reliability and other NQF criteria.

Michael Lieberman: That's what we're trying to account for the issue of, you know, and one of the examples of via checkbox for meds reconciliation and certainly it's the ability to capture that data is highly feasible and pretty universal. But in terms of whether or not a real life, clinical – clinically accurate medication reconciliation occurred, we feel like – and it's been an assessment of whether that data reflects clinical accuracy as opposed to someone just checking the box.

Male: Right. We clearly agree but that is actually what (are) validity, you know, the data element validity criteria build with which is outside of our feasibility assessment when we're just generally assessing measured in general.

Female: I think it's fair to (kind) of that statement about where we think it's reasonable to assess the feasibility of accuracy based on current workflows.

Male: (Chris), I just want to add ...

Male: Again, I thought that we cannot talk about – understanding that this – that we were double counting validity or the validity was addressed elsewhere but we felt that this assessment was supposed to be done, you know, very early on and whereas validity may come later in the – in the process. So, this would help to kind of uncover that issue or to include that issue very early in the process as well.

Male: It gives an eMeasure perspective on validity. So, oftentimes the CSAC doesn't necessary know when this may be valid but this is really captured in accurate, you know, available format. So, I think – we're all sense of that is that – there's some overlapped with all these things but we are at trying to do some value-add. These are upfront for as they – an eMeasure perspective to the validity as it is.

Rute Martins: This is Rute. I agree that there is some overlapped but I wouldn't even call it overlap. I think this – from the perspective of feasibilities looking at the accuracy of data, we're not looking at validity or accuracy. We can't. This is – we don't – we won't have the data to do that at the stage that where we're doing feasibility or at the multiple stages, I guess we would be looking at the potential for validity and accuracy.

So, it's more of a preamble to validity and accuracy and kind of a way to stir the development process in a different direction if we find something that is simply not – we know that it's not going to be reliable. There's a high likelihood that it's not going to be reliable based on basically feedback from vendors and providers.

Michael Lieberman: All right. So, can we move on? Next comment, line 28.

Shannon Sims: This is Shannon. I would say that we certainly would agree with that as a group. I'm not sure how it plays into feasibility. (I think) – just we move on to the next comment.

Michael Lieberman: OK. I agree. All right, line 29.

Shannon Sims: This is Shannon, again. I think I would have a similar statement. I'm not sure I agree that mapping is typically – is actually a simple task. (Mean) it will

certainly supportive of standards, but again, I don't say it was placed into feasibility. And we've discussed workflow, I've (inaudible) so I'm not sure this one ...

Michael Lieberman: Yes. All right. So agree, move on? Largely, OK. So, next on line 30. And on the – OK, (we'll just) hold there.

Howard Bregman: I do think the commenter is correct by referring to the fact that we are inconsistent in the document about whether we referred to a single EHR or multiple EHRs. The goal of our assessment I think we all agree is to get represented scores from multiple EHR systems, but for any single assessment done by a single provider or vendor. It's only going to reflect one for the most part.

So, the – any given score is really going to, again, reflect one EHR vendor or one provider situation but it's the combination of multiple assessments for single element that's going to result in scores across vendors and across providers.

Michael Lieberman: Ah.

Howard Bregman: I think that's confusing in the document in some cases.

Reva Winkler: Howard, this Reva. I'm not totally clear on what you're trying to say because the recommendations include, you know, assessment with multiple vendors or of systems. We'd talked about that last time. How do we make it more clear?

Howard Bregman: I have to identify – sorry, I don't have – I'm not prepared to say which areas. I know the – which areas of the document need to be changed but I could go through it to identify those but this commenter talks about page 12, and just to look where she is ...

Aldo Tinoco: This is Aldo. I think that I don't entirely agree with the comment. I'm familiar with the reliability across EHR systems or across implementation but I see that more as the output or the calculative results of the measure as to whether or not the algorithm performed as expected against some standard like a validation deck, but the ability to actually program or implement a

logic, I think is an (act of) feasibility. We want to look at that across different systems. It's a new ones but I think they're different.

Rute Martins: And this is Rute, and I agree. I guess the question is when is it feasible for us to do that kind of feasibility assessment because once you have a piece of logic for a particular criterion, for example, gestational age and timing around gestational age, so some of the feedbacks that we've had from vendors is that that timing piece around that – a particular data elements may (just) whether or not the data element is feasible but the context is important.

So, in that regard, I do agree with you Aldo. We need to make sure that we're assessing the feasibility and the right context. I would also say, though that actually verifying that the measure computes exactly the same way across EHRs. It may be too much – too much work put into the feasibility piece of this and it actually transpires a little bit to the validity and reliability piece.

So, I guess, what I'm saying is we probably should be clearer about what we mean by feasibility of the logic and how far we are expecting measure developers to go in assessing the feasibility of the logic.

Male: Yes, I would agree. I think it's really difficult to say by looking at the logic thing, you know, this is either feasible or not feasible to implement across multiple EHRs. You know, I'm not as that as familiar with the measuring authoring tools as others on this call but it would seem, you know, if the – if the logic can be expressed in the measure authoring tool wouldn't that kind of make it feasible for implementation across EHRs?

Rute Martins: This is Rute. And no, that's the problem is that you can express – you can actually create unfeasible logic pattern. It's – sometimes the timing relationships that you're requiring in the measure are not reasonable within the workflow. Something of that nature, for instance, and again, the timing of when you're looking to capture those data elements, so let's say that you have a data element that's gestational age and that it's actually captured and in EHRs in structured formats and that would be fine.

But in your measure, you're looking for the gestational age other documented on the day of delivery for instance and most of the EHRs allow for the

documentation of gestational age and the workflow that is they document the gestational age when the patient is admitted.

And so then you have a disconnect right there that they actually don't have the data element within the time frame that you're looking for it, and that creates a feasibility issue as a result of the logic but not directly a data element of in the value sense.

Male: I don't know if I would call that logic though or is that – is that feasibility if that's where we're getting at or if it's more – if that has to do with the workflow of that data element because you're saying basically that your data it's not captured at the time you need it.

Shannon Sims: This is Shannon. I'd agree with that (meaning).

Rute Martins: So, then I think we need to clarify what we mean by data element because as far as I understood it, we were talking about QDM element. So, it's going to be a category. It's going to be a QDM data type attached to a valued set and potentially have them attribute.

When I – when I extend – when we say logic – feasibility of the logic to me that goes a little bit beyond that. It goes with the timing around that QDM elements and it may involve more than one QDM element and it would be kind of a bullet in the human readable of the eMeasures.

So, I guess the individual criterion or criteria for an eMeasure and that's the – to me that's how I see we need to evaluate in this feasibility piece the logic. I don't see that the connection of “ands and ors” and all of that belonged in them, the feasibility piece but assessing the feasibility of individual criteria is different than assessing the feasibility of a particular QDM element.

Michael Lieberman: Yes, I would. I mean, I think you bring up a very good point and that's what I would – that I was thinking about more would be kind of the context of the data element within the measure might affect its feasibility. So, we have talked about how we want to have a library of data elements with (pre-computed) feasibility but that's probably not sufficient because, you know, the example you just gave, it might be that a data element scores one way but

there's something in particular about the way the measure is using it that makes it less feasible.

And, you know, I don't know how to actually formally capture that in this – in this assessment other than the, you know, perhaps have a, you know, separate section about context specific feasibility for a data element or that you would actually go and change the standard numbers in this – for this case, but then getting back – that's what I – that's why I would kind of address both the context issue and that's where – and I can – I think you said the same thing about the logic when we're really talking about the “ands and ors” and whether or not it can be, you know, it cannot be computed.

The – trying to say, we – I don't think we have that – a formal way to evaluating that between these processes. I think with the comment that was saying that's part of feasibility in general. Maybe either we should live that out or, you know, point out that we might want to develop some sort of – there may be some sort of way of coming up with a metric just around, you know, the complexity of that logic.

Fred Rachman: This is Fred. And I think – this is with the overlap between the other criteria really – (what it makes) confusing for me, just looking from other commenter frames this and then when they (quote) that, you know, that the calculates (inaudible) with the work for multiple – within multiple EHR systems.

To me, that sounds like they're talking about – you end up with measured scores that you could compare across EHR system and that's ...

Michael Lieberman: Yes.

Fred Rachman: ... about reliability than it is about the feasibility even gets this.

Michael Lieberman: Yes. Now, I would agree. So, I don't know if we want – do we – would we agree with talking out the kind of assessment of measured logic as part of the feasibility or is there – are there any other (super) recommendations about what do with that?

Male: Could we add one or more EHRs?

Michael Lieberman: Well, I think that – I think we already kind of address that by talking about assessing through multiple EHRs and multiple implementations when possible.

Male: But I think the specific comment is sort of addressing what (Chris) that logic really needs to work in context of a – whatever even single EHR implementation because that's sort of a priority determines whether it is taking into consideration the proper context and so forth but that the ability for it to be reliably reproduced in multiple EHRs is kind of I guess under a different criterion.

Howard Bregman: Can I suggest that I – I think (what) we really need to do is in this document, make it clear that the idea here is that when a measure developer is working on a measure they're going to take the form on page 10 and 11. They're going to send it to multiple people, multiple organizations as many as they can get to participate.

Each of those contributors that's going to fill out the form is going to do it just based on their own situation and the information that they have about their own whether it's vendor or how they use in the EHR system or otherwise and so each of those inputs that they're going to get is going to represent one. They're going to send all these copies of these forms back to the developer. The developers going to compile them all and then come up with a conclusion and the conclusion is going to be ideally based on what we've recommended an X number of providers, Y number of vendors represented and then they're going to make a conclusion based on that.

And I think if we can just make that clear, that would be fine, it's just that our language sometimes when if on the form itself it says, "You're talking about majority of EHRs, which it does on page 10, it's confusing because really all you're really going to represent for the most part is one EHR."

Male: Yes, and I agree with Howard. I think – I think that sort of the concern that's being expressed even though that's not what we mean. It probably is not clear to them.

Rute Martins: This is Rute. This is actually a comment that I've made before and I do feel strongly about this and I know that some of you don't that I do think that the tool should work for a single EHR and right now the way that the scale is set-up, it doesn't allow us to evaluate a single EHR independently and then aggregate results. If an aggregation in itself and I think that maybe were part of the confusion comes from.

Michael Lieberman: I mean, that's an interesting point. So, to kind of rewrite the form from the – from the point of view of a person evaluating any implementation of an EHR., so when you look at – well, I mean, it is still – data – so, I think that would be mostly be under data availability. I think data standards, workflow and data accuracy can all be addressed with an implementation.

Rute, where do you see the issue with not having this be applicable to one system?

Rute Martins: I can remember what the criterion is exactly but you just mentioned the example the majority of EHRs have it. So, maybe data availability, but when you – when the scale is the majority of EHRs and you're already aggregating across EHRs.

Michael Lieberman: Yes.

Rute Martins: You're answering the criterion. I can't really pinpoint it. But I ...

Michael Lieberman: OK.

Rute Martins: ... you know, that some of the – part of the scale is not – cannot be applicable to a specific EHR.

Male: So, isn't the tool both for measured developers to ask themselves and we could have certainly special version that gets distributed to individual EHR vendors or individual proprietors to use a single EHR? Is that a problem?

Michael Lieberman: Yes, you know, I think that's a good idea. I think that's exactly right, just, to kind of modify the language in the form for purposes of evaluation of a single implementation.

Male Right and then provide guidance for, you know, when you get all these like Howard was describing, the developer or whoever is doing the assessment then they would need at least, you know, X number of those, hopefully, more – the more the better and then that would be the sort of way to score in aggregate at that point.

Michael Lieberman: Yes, and that might, you know, we've kind of – so, you know, another idea – another thing to do would be to actually include in this document a propose or a possible workflow for evaluating a measure overall which is what we said and I think we've kind of gone around the idea of who's actually going to be doing this and who's going to be doing the work but if we say, you know, one example would a measure developer would send us out to EHR, whatever that might be but could give kind of to just specifically go through the steps of how a complete measure might be evaluated for feasibility use in this tool. It could be – and it could be a supplement for the document or something as well.

Male: Which actually would address the first part of the comment that, you know, gave a specific guidance how we would use the tools.

Michael Lieberman: Yes.

Female: Is there somebody who wants to take a shot at drafting something like that from the committee?

Michael Lieberman: Without volunteering anybody but I – it would seem like a measure developer has most of the experience with this process and if they would – if one of – one of the measure developers wanted to kind of just go through a step-by-step process of how you would do it using this tool, that might be – that would be really helpful.

Female: And Keri said – and I don't want to volunteer Keri for this but Keri, we were talking about going through some of the – some measures that you are developing right now. I could – I could certainly participate as well and we could use those as a reference point to apply the tool, too.

Michael Lieberman: Yes, and then you can actually include it as an example of how the – how the tool could be used as opposed to a propose, you know, you don't want to be so prescriptive of telling people exactly how to do it but you want to show them how it can be used.

Female: Does that sound good to you, Keri?

Reva Winkler: Keri, still on the line?

Female: Not.

Michael Lieberman: Maybe not.

Female: Well, she did – she did offer to work offline. So, I think we can connect offline and then – and see where we can go from here.

Michael Lieberman: OK. That'd be very good. OK. And then, so – it was (told) – they do mention this one specific line and before we move on, so we have under – on page 12 of the draft report, it has an additional to the scores for the individual data elements, eMeasure feasibility assessment report should include, and then it has assessment of the feasibility of aggregate data elements and some sub-bullets. But then it also says assessment of the measure logic and then it has sub-bullets – does the calculation algorithm work in multiple EHR systems; how complex is the logic; how easy is it to explain to providers.

And I think the latter too (inaudible) another comment about, how complex is the logic and how easy is it to explain to the providers, is something that I think is kind of within the scope of this. I think, does the calculation algorithm work in multiple EHRs, I think it was although mentioned that that's really a more of a validity or reliability issue once the measure is out there. And you could – you could calculate, you know – how likely is the calculation algorithm to work in multiple EHR systems. But I don't know if that, you know – if that's again something that needs to be done during kind of a feasibility stage or later.

Female: Do you want to remove it?

Michael Lieberman: Yes. I would just say remove the, "Does the calculation algorithm work in multiple EHR systems." Does anybody else either agree or disagree?

Rute Martins: This is Rute. I agree.

Female: I'm OK with that.

Michael Lieberman: OK. All right, then moving on to the next comment on – I think that's line 31 that we're on now?

I think the we are in agreement with that and that's the idea behind this is to use as the method of exposing of issues early on so that you can discuss alternate data on alternate measures.

Female: I would agree.

Male: Same.

Michael Lieberman: OK. Next, line 32?

Male: I just think this comment doesn't fall under our (umbrella).

Michael Lieberman: Yes.

Male: Particularly about measure development.

Male: I agree.

Male: Yes.

Michael Lieberman: I agree as well.

Female: Yes.

Michael Lieberman: OK. Line 33?

Howard Bregman: I think issue of the specification of measures is an important issue, but it's not really a feasibility issue.

Michael Lieberman: Yes, and we actually do address a little bit of that, you know, assessment of measure logic. We did include how complex is the logic and how easy is it to explain to providers, you know, part of that kind of initial feasibility, which would address some of these issues.

Male: Mike, do you know where that is, where we say that?

Michael Lieberman: Yes, on page 12 of the document under assessment of the measure logic towards the end of recommendation, too.

Male: Do you recall why we put that in?

Michael Lieberman: Well, I think we want to get – it suppresses by in addition to the scores for the individual data elements, eMeasure feasibility assessment report should include, and then it has assessment of the feasibility of the aggregate data element and some sub-bullets there, and then assessment of the measure logic. So, I think what we were trying to say is that, you know, assessment of each individual data element is not sufficient to assess the feasibility of the overall measure, because there are other aspects to it.

Howard Bregman: I do think it's important to look at feasibility in context to the measure and it's also important to say whether the measure as constructed will collect the data that is intended to be collected for the purposes on the measure. I'm not sure if really complexity of the logic or how easy this is explained to providers is really something that's part – should be part of the eMeasure feasibility process looking at it again. I wonder if anyone else asks.

Rute Martins: Yes. And Howard, this is – I actually agree. Because if you think about and this is something that we had discussed before, you know, what is the complexity of the logic? Is that we have seven pages of ands and or that that's burdensome? Should that be counted under feasibility? And then the other aspect of it how easy is it to explain the logic to providers, because it may be really easy to explain the intent of the measure to a provider and they may want to understand every single aspect of the logic, but the logic is driven by constructs that sometimes are beyond the scope of what a provider is very familiar with, such as (Gideon) constructs and (angel seven) timing relationships and all of that.

And so, part of the talent representing appropriate logic sometimes it doesn't create complexity in order to accommodate for different representations of data in EHR, for instance, different ways data can be recorded in the EHR and it does make the logic more complex. But will it make the measure infeasible? It may actually make it more feasible because it may fit more work flows, for instance.

Howard Bregman: Are you suggesting we remove those items from that reference on page 12?

Rute Martins: I think it's very important to address what you – what you just said, the data elements in the context of the measure, and so pieces of the logic may be important to assess. I would say that it's probably beyond the scope of what we're trying to do with feasibility to address how many pages the logic has, how many data elements, does the logic address if they're all feasible, it doesn't really matter.

Howard Bregman: Yes, I have to agree with that.

Michael Lieberman: Are there other – are there other points in the measure developing process fairly early on where these are – where these are addressed? So I think that kind of the complexity and the ability to explain it to providers is getting at an important issue that when you – once you implement a measure it's so – it's so complex and kind of in the logic, it's so concluded that it's tough to – and you can't explain it to a provider then you do run into the issues where they don't – they don't want to use that and they have trouble accessing it. And they don't like the idea of the black box.

Female: (Inaudible).

Michael Lieberman: Go.

Rute Martins: This is Rute again. I guess – I guess my challenge was that – is that it gets into EHR construct, QDM data type. Is it really reasonable that all providers are going to have to understand all of the aspects of every single piece of logic in order to accept the measure and where is the trade off? Do we make a measure less exact in order for them to be able to understand it quickly?

It may be understandable, it may not be easy to understand, but we have that same challenge with paper-based measures. Sometimes the logic is complex and the number of layers are they're many. And so, I guess the question is does that make the measure less feasible in terms of the data collection and data aggregation and calculation, it really doesn't. It may speak to other NQF endorsement criteria, such as importance and so the measure may be perceived as less important if it's not very understandable. But I don't think it speaks the feasibility.

Michael Lieberman: Anybody else? So, I think one – the recommendation then will be to remove the complete section on page 12 about assessment of the measure logic. I mean, there are three bullets – there's a bullet point and three sub-bullets. Does anybody else feel strongly ...

Howard Bregman: I do – I do, Mike, think that we do want to say that part of the feasibility assessment is assessing the elements in the context of which they're used in it in the measure, not just as individual – not just as completely independent entities that are not being looked at this context – in context.

Michael Lieberman: OK.

Howard Bregman: So, I do think that there should be a statement about assessing the data elements in the context of the measure logic. I think Rute and my points are really about whether the assessment should also address the complex in a logic or how easy it is to explain to the providers, because I think she explained well that those things really don't seem to fall under the umbrella of eMeasure feasibility.

Rute Martins: And I would completely concur. I do think it's very important to assess the data elements in the context of the logic. And just as an example, I think it may be using the same tool that we used, let's say, that we're doing the first – our first pass on feasibility of looking at the capture of certain data elements.

If the data elements aren't captured at all, do we go different route or do we try and get a little bit more detail around that data element and do feasibility check again once we have the logic around that particular data element? And

then we can actually see if the feasibility issue is with the data element itself as a QDM data element or is it the context in which it's being used in the measure that's the problem. I think that's actually very valuable in terms of feasibility.

Female: I agree completely to what you just said. I think we talked about this quite a bit at our in-person meeting as well.

Michael Lieberman: OK? So, it sounds like every – those who have spoken up been in favor of removing the ones around logic and adding in a comment about context of the measure – of the data only within the measure, anybody disagree?

OK, we'll move on then. So, line number 34?

Yes, so the first one I think we have addressed, if we're going to remove it.

Howard Bregman: The commenter says data collection burden traditionally refers to the burden on healthcare providers themselves. In this report, it appears the burden on EHR developers is a factor in determining feasibility and it also says measure developers. I think that our intention is that the burden is really – is referring to the providers. I don't think I'm not sure the specific language, but I don't think it was meant to refer to burden on vendors or measure developers.

Rute Martins: This is Rute. I actually think that part of the framework does speak a little bit to the burden on EHR developer is, for instance, there are no structured fields within EHR to capture particular data element that will require effort not only on the providers side in order to capture that data element infrastructure but also on potentially on the EHR vendor side.

What I think that the commoner is confusing is our discussion around the burden of completing the feasibility assessment, which is actually a burden on the developer, as we discuss this in the report versus what we're trying to measure with the assessment, which is burden on the provider. So, I think there may be some confusion here on the part of the commenter. I don't know how our language in the report may be causing this, but it seems to be – it seems to me that we may be using the word feasibility on the measure

developer side of this feasibility of the process of assessing feasibility, so that's why it may be confusing.

Zahid Butt: This is Zahid. I think that's a different angle, but I think that even if you look at the providers and the burden on them, the question isn't the context of feasibility. Shouldn't it also include the potential for having to develop new elements and so forth if part of feasibility might address some of that? That clearly it is a provider burden at the end of the day for most data capture if not all, but there's an element of the vendors in there to the extent that something is not feasible because no one has it built into their systems.

Michael Lieberman: Yes. And this commenter is specifically referring to the first overarching principle on page 7, where we have this second; EHR vendors indicate that all data elements could be ultimately integrated into an EHR. However, the questions are whether the importance of the measure results justifies the cost and the time required for development and whether clinical workflow can efficiently capture the necessary data as a byproduct of the provision and documentation of patient care. Until agreement is reached that everyone must achieve the same level of interconnection and maturity, feasibility will be variable across all EHR implementations.

So, I think what they're getting at is it makes a sound like, you know, part – there were addressing this issue of the importance of a measure justifies the cost and time required for development and whether clinical workflow can efficiently capture. So, what is the purpose of having that that statement in there?

Zahid Butt: So, Mike, from – this is Zahid again. I think from what you just read I think it captures it accurately, because it addresses the provider burden and it also addresses potentially the burden which, you know, in aggregate would have to be taken into account if new elements would have to be implemented or incorporated within systems. So, I am – I'm happy with the statement that you read, which is in the report because it does address both issues, which I think are equally important or certainly the developer side is also important. But certainly, I'll be interested and hearing what others have to say.

Howard Bregman: I think our language is OK. I do think that part of the assessment is assessing the burden on the provider and the overall feasibility reflects somewhat the responsibility – what the vendor will have to do to make it – to make it work.

Michael Lieberman: So, I think then – so we're saying that most of the time we talk about burden of capture, we're talking about provider. And I don't know, do we need to be more explicit about that or not? But I'm happy with the way it is now.

Rute Martins: This is Rute. I think so, too. I think this may be some confusion around the use of (word) feasibility but if taken in context that I think we're fine.

Michael Lieberman: OK. OK, next comment, line 35? The measure specification and calculation logic – the measure specifications and the calculation logic are important to understanding the intent of the measure (inaudible) on the measure development.

So, I'm not sure which – is that – is that back to what we've gotten Rute or not? No, yes?

Rute Martins: Doesn't it go back to the logic evaluation?

Michael Lieberman: That's what I'm trying to figure out. I don't know exactly which second sentence – oh, I see that the second sentence the measure specification and calculation logic. Oh, the measure – so it says, the TEP emphasized that feasibility is not solely about the data elements. The measure specifications and the calculation logic are important to understanding the intent of the measure which can influence what data must be collected.

The number of data element – so I think that there – although the measure specification and the calculation logic are important understanding the intent measure may not be present early on in measure development. So, they won't – I would assume that on page 11 on the first – it's actually the first – the second sentence of 2.2 eMeasure Feasibility Assessment.

Rute Martins: Mike, I think this goes back to the idea of the criteria and the context of the data element in the logic ...

Michael Lieberman: Yes.

Rute Martins: ... more a full blown set of specification and calculation algorithm. So, I think it speaks to that duality of where the logic is important to assess feasibility versus where it's not. I believe that the commenter is speaking to idea, well, we don't – we won't have full measure specifications when we talk for feasibility. And I agree, but I think with the modifications that we've made to the bullets on the logic and if we clarify that, we are really intending to assess the feasibility of a data element within the context of the specific logic criterion, for instance, that will help clarify.

Michael Lieberman: OK. Anybody else or can we move on?

Let's move on to line 36 now. (Inaudible) – so that one is in agreement so we can move on.

Line 37? So, I think the comment actually has two things as well. So, it says no measure should be considered feasible and not all data elements in the related context is feasible. And I guess, you know, I would go back to our kind of overriding comment that we're not – that we're not looking for a yes-no, feasible-not feasible. We're looking for level of feasibilities, so that's where we're kind of addressed that that part of the comment.

Second part, measure feasibility regarding the ability of the logic to score measure correctly is highly significant, but in different issues and data feasibility. These statements seem to combine it too. So, I think we've kind of gotten the way from looking at logic so I think we can – I think we kind of addressed that as well.

Anybody else have a comment there?

Howard Bregman: I will also so say that we have made the comment before some have made the comment that a measure is only as feasible as at least feasible data element. But it's not quite true, because an exclusion data element is much less valuable than a primary data element, such as the data element which captures the treatment that is to be given. So, we shouldn't be using if we are or we

shouldn't be saying that statement, which is that a measure is only as feasible as at least feasible data element.

Michael Lieberman: I would agree.

Rute Martins: And this is Rute. I agree too and actually it's something that we've been thinking about of what in terms of if it's specific, if a criterion is included in the measure that covers a very specific set of hospitals, let's say that only academic hospitals do this one procedure that it is an exclusion for measure and it's not particularly feasible. Does the fact that the burden is on a particular set of providers that is small and that presented may be a very feasible data element but for other providers that may not be?

It's interesting to see how a data element, depending on how it's used in the measure and the extent of coverage that it has significance in the overall rates or particular provider can influence or not the feasibility of the measure. So, just because the criterion is there, it doesn't seem that the measure can't be implemented even though the particular criterion isn't there feasible, so I would agree with Howard's comment.

Michael Lieberman: Yes. That's a great point. So in our – I'm not sure that we out in a way of formally capturing that. So, if you have an exclusion data element that we would say that score is low in the feasibility across all implementations yet is really only applicable to certain areas and in those areas it's captured reliably. That actually you could include that data element and it could be – and overall the measure could be very feasible yet we'd still have that one low feasibility score.

Howard Bregman: Well, I would even go further than that and say, "Supposed you have an exclusion that's really only going to affect 2 percent of your population, you would expect to only affect 2 percent." And supposed that element is completely infeasible everywhere, Mike, I still think that measure should be rated fairly high feasibility. And you know what? You'll have to either forget about the 2 percent and just record them as a measure failure or work around it in some way.

Michael Lieberman: Well, what I would assume is that if you have a specific population that where your population is much more than 2 percent that you would figure out a way to capture that data. If it's going to – I mean, if it's 50 percent of your patients and you want to make sure that they're excluded, it's going to be much more important of you to figure that out ...

Howard Bregman: Right.

Michael Lieberman: ... where everybody else might not need to.

Paul Tang: Let me try to – this is Paul. Let me try to drill down on this and understand this better. In a sense, isn't the better approach to not have the exclusions for the 2 percent than to include it and then have everybody look for it? I'm not sure that – I think when we published measure spec and we published an exclusion criteria then in a sense we're looking for people to all look for that exclusion. That's one way of interpreting it.

Are you saying from measure development point of view, you're looking at it in a different way? In other words, people can pay attention to the summary exclusions if they choose to.

Rute Martins: This is Rute. I wouldn't say that. But let's say that you're talking about a critical access hospital and you're 99 percent of the time the patient that comes in will be not be on a clinical trial and this is from an inpatient perspective. So, from an outpatient perspective, I think it's more difficult I would say. But in that hospital, you can either create a form where for every single patient, clinician has to say whether they're not – they're clinical trial or not, you can implement a workflow where you allow for a clinician to say of a patient is in a clinical trial only when they are, you know, because there are different ways of forcing certain workflows.

Or if that particular hospital knew that they are not going to get patients in clinical trials and they just don't want to capture it in the structure format, then they can assume that they will not be credited on that exclusion. But it provides the flexibility to the hospital and I think it does. I mean, the measure specification – they do from a paper perspective, you do have to answer every single data element.

From an EHR perspective, some of these exclusions may be catch all exclusions that are not applicable to a particular hospital. And I don't think there's a final answer on how this shall be implemented across all providers. But certainly, there is a possibility that a hospital chooses, for instance, not to document reasons for not in a structure format and they realized that this is – this is a mechanism that would allow them to pass the measure in exceptional situation, but they're not willing to do the work and so they agree with the rate. And then, you help to convert situation which the hospital who wants very badly to reach that 100 percent and that is only feasible if you do have you do have the possibility of expressing exceptional situations.

I think there's certainly a balance between what you are representing in the measure, but I also think there's this belief that every single criterion that is in the measure specification has to be explicitly documented for every case. But the measure really is looking for the information if it's there, if it's not. If it's not there then you're going to fail that particular criterion wherever it takes you. So, there's no longer that requirement that you shall touch every single data element in order to produce the measure rate. You may have a lower rate that may or may not be accurate, but that will depend on how much your particular provider is willing to go to capture that information.

Paul Tang: Did you say that on paper you are expected to touch every data element in a sense?

Rute Martins: Yes, because in the paper-based data collection, the data elements are not necessarily – they're not necessarily derivable from a structured field in the medical record and they require that the abstractor see whether – the abstractor has to say if a patient is in clinical trial or not. What the abstractor does, they go through the record and they see if there's no documentation of the patient being in a clinical trial they will answer no ...

Paul Tang: OK.

Rute Martins: ... but they will have to answer no. So, the measure speaks.

Paul Tang: I understand that. So, I'm wondering how people are instructed to act differently or that the definition actually means something very different when you describe an eMeasure versus one that's implemented on paper, right? I completely appreciate that.

Male: (Inaudible).

Zahid Butt: This is Zahid. Just to add a little bit more to that. I think there are a couple of issues there. One, like in the example who give about the clinical trials, we have many instances in which the examples you gave that they never ever do clinical trials. And so the question is answered but it is default to no for every case, so in that sense, you know, you can sort of answer the question but not necessarily you have to check every case necessarily, so that's one aspect.

The other is that there is something called skip logic within the paper-based measures, so that you don't have to answer every question in every case, so that if you reach a certain point where the rest of the questions are not relevant, you could choose not to answer those questions.

Paul Tang: But who said that if you, let's say, you did – you do clinical trials on a very, very small number of patient and you could just choose to take that, in other words your denominator would include folks that are on clinical trial and so your number would be lower. And that – so in the sense, unlike the paper you are no longer truly attesting to the fact that the fact that this person on a clinical trial had no mention of them being on one is OK.

Howard Bregman: So, Paul, in that case, where you have really small numbers it obviously is going to negatively impact your rate because the case will be in the denominator.

Paul Tang: Right.

Howard Bregman: But it would fail because, you know, the patient was not in trial so, you know. That's a decision that folks will have to make. Generally speaking, when you have small numbers of cases in the denominator, they are very sensitive to even one or two cases falling out because it would affect your rate a lot. So, generally, not to ignore that.

Rute Martins: Paul ...

Michael Lieberman: And I think that's the problem we run into if we start trying to say that, you know, people can kind of ignore the exclusion. When you look at where, you know, we're working on core measures and where, you know, the 50th percentile score is 98 percent, you can't ignore even small numbers it might fall out and if these measures get you more and more performance programs, you program that not one even let go of, you know – the possibility of a case or two.

Paul Tang: Yes, and ...

Rute Martins: Paul, you know, I think this is a very slippery (inaudible). I'm not saying that provider should be ignoring exclusion, but I do think that there is this belief and I don't know really how to put in simple terms. I don't think it's the ignoring the exclusions, it's the other way. So if we – if we specify measure that makes the space for all of this flexibility so that we have – or let's put it in other way.

Let's say that the clinical trial is very infeasible. We can include it in the measures facts and everyone is going to be screening and kicking because we included that it is not feasible. Yet, if we don't include it, we're already ignoring it at the measure specs level. So, it means to have different effects on different kinds of providers and it's certainly will and that will become a problem of its own.

Paul Tang: Right. So, let me – would I if I could just lay out this the assumptions that I have before this conversation took place and see where either I was the missed directive or whether the field, the community may have share similar understanding. So, I thought a little bit like what you described in the paper world that you had to fill up all of the data element, including exclusion so that you could reliably and accurately report on your numerator and denominator. In the sense, when you fill up a paper, you're attesting to all these things as fact.

I did not think that there was a different situation in the eMeasure from that point of view, and that is one of the reasons why I thought from an NQF point of view, we really did want measure developers to deliberately consider each and every exclusion to make sure that really was material and that the cost outweighed – the benefit outweighed the cost and do it at that level, i.e. the measure definition level, not have every provider decide whether it's worth in capturing the data elements for exclusion.

So, this is very new to me, and I want to make sure that I've now understood a new norm that I don't think it's rightly understood.

Female:

And Paul, what I'm – I'm not saying that providers should be choosing exclusions to implement or not implement, but there – and I think what you've just described is a very clear distinction between the paper-based world and the eMeasure world, and it can be found in a sentence, which is the paper-based measures make a distinction between data elements that are not touched by an abstractor. So, we really don't know the answer to the data element, and we call that missing data.

And then, there is information that you can't get from the record because you can't determine the data because there is two conflicting dates, let's say. And we call that unable to determine information. From an eMeasure perspective, you can't distinguish missing data from failing a criteria. So, for instance, if you are an in-depth missing data in eMeasures includes the unable to determine.

So, as an abstractor, you can't find the information in the record regarding new trends at the clinical trials. You will answer no to the data element clinical trial, and that would be considered in the eMeasure logic. If you fail to answer that data element, that whole case will be, or maybe thrown out because the data is missing, and so you can't calculate the measure.

From an eMeasure perspective, that is no longer the case. So, if the data is missing because you don't have that attestation stuff, if the data is missing, then you will fail that criterion, and the measure will still be calculated. So, there's not – there's not a missing data policy around eMeasures as there is

around paper-based measures. And the kind of missing data policy that we can put around eMeasures is going to have to look because there is no – there is that attestation stuff.

Reva Winkler: Mike, this is Reva. I just wanted to note we've got limited time left.

Female: Yes, and ...

Paul Tang: Do you think that clarification we've just made is known by everybody?

Michael Lieberman: Well, Paul, you should know that in a meeting with CMS and ONC, the EHRA raised this issue this week. It was not the first time we raised the issue.

Female: Right.

Michael Lieberman: Apparently, both (Jim Walker) and (John Halamka) feel that the intention of the policy committee was that measure exclusion should be optional based on the desire of the provider.

Female: Yes, and – I'm sorry.

Michael Lieberman: And this was raised in the meeting that we had with CMS and ONC. I don't think the question was answered definitively, but it was certainly discussed.

Paul Tang: So, I'm kind of take this offline. I just – I'm amazed, I'm totally caught by surprise ...

Female: Yes, and (inaudible) also. I mean, I think you have to make exclusive policy ...

Female: Yes.

Female: ... about missing data no matter what. If it's missing EHR or if the provider elects not to answer it, then there has to be some policy about how that's going to look in the measure.

Helen Burstin: Yes. And actually – this is Helen. I agree. This is a really important issue and probably is bigger than specifically the eMeasure feasibility discussion we're having now, but you know, (inaudible) methodology actually and I had a conversation with Kevin Larsen this week about this exact topic. And we have to have consistency in the way exclusions are done. If they're optional, then you'll wind up with apples – you know, then you'll wind up with data that cannot have valid comparison.

So, I think there's a distinction as well between exceptions and exclusions. I mean, if it's truly an exclusion, and it's you know, it clearly is backed up by evidence, and without it, the results are distorted. It just needs to be there consistently. But I think you've raised an important point, Paul and others that we need to come back to.

Paul Tang: Yes. So, I'm happy if there is a way we can arrange this, and fortunately I do think it impacts feasibility, that's for sure.

Helen Burstin: It does, yes.

Paul Tang: Happy to like participate in some other call that doesn't take this time, but I think it's eMeasure.

Helen Burstin: Yes, Paul, I'll keep you on the loop on the – on the Kevin-Karen conversation.

Paul Tang: Yes, thank you.

Zahid Butt: This is Zahid. Just for the record, I'll just throw in one more thing that missing data policy in the abstracted measures is one of the biggest headaches that we have to deal with in terms of how to deal with data that's not there because it tries to distinguish between missing data versus not being done. It's a huge headache. So, the more we can get rid of that, the better it would be.

Rute Martins: And this is Rute. I just want to make final comment on this. It's that the eMeasures really deal with missing data. The way the logic is setup here, criteria is going to be true or false, basing on – based on whether the data is in the EHR or not. The way you implement each exclusion in each data elements may provide different burden and different accuracy to the data that

you're capturing, and I do think that the feasibility framework that we have right now does account for that.

So, if you have the field there populated often, or is it not. And if it's not populated, what does that mean? Does it mean that it isn't reliable, or does it mean that you're only looking for the smoking gun such as a clinical trial? It's only populated when there is, in fact, that inclusion for instance.

Michael Lieberman: Yes. All right. Let's see if we can go get to a few more here.

Reva Winkler: Mike, it's Reva. We've finished all of the comments on the recommendations around the scoring, and the next two on the comments and recommendation three and four, probably not a tight priority. I'd like to skip down to line 46, where we begin the recommendations for recommendations five, which is around NQF's evaluation for endorsement, because that's something I really want to hear the committee's thoughts on the comments.

Michael Lieberman: OK. So, line 46?

Reva Winkler: Yes.

Michael Lieberman: OK. Let's read that.

Reva Winkler: I think this first one was one that came up with some of the discussion previously.

Michael Lieberman: Yes. I think we – didn't we come up with new wordings for that, Reva?

Reva Winkler: Yes. We just want to use multiple systems in no specific installations.

Michael Lieberman: OK. Let's move on to number 47. So, first of all, should the scorecard be optional or should it be required? Do we state that it's required at this point or not?

Reva Winkler: It's implied in that the criteria for considering an eMeasure would include the results of the eMeasure feasibility and assessment be included in the measure submission for consideration.

Michael Lieberman: OK. I mean, so the alternative would be, you know, strongly recommend the inclusion and leave it up to NQF to decide whether a submission without the scorecard should be considered.

Reva Winkler: Well, I'm wondering, would a reasonable response be either use of the scorecard or a summary of a fully-transparent equivalent information.

Rute Martins: This is Rute, and I think that's a great idea, the equivalent thing. I don't think we want to be too prescriptive, because we do want to – we said multiple times that we want this framework to evolve, and that will only be possible if measure developers are developing either other scorecard or additions to what we have here.

Michael Lieberman: OK. Anybody else want to either agree or disagree with allowing comparable materials as an alternative?

Ginny Meadows: This is Ginny. I agree.

Michael Lieberman: And then, I mean, that does – Reva, so that put some burden on the NQF to determine whether something is comparable.

Reva Winkler: Correct.

Michael Lieberman: OK. OK, next comment, or actually second part of that comment, there was, we further recommend removing requirement to measure developers including plans for improving data on it. OK. So, page 13, so may I guess that I kind of agree with the comment. I think the – what we wanted, you know, is you could always – you could require a comment on why it's scored to one and you know, what might be done.

But I don't know that we – we want to let them off the hook completely and just say, you know, that's kind of beyond the scope of the measure developer.

Reva Winkler: This is Reva. My question would be, is if indeed the developer has identified feasibility concerns with their measure, what's their response to having that problem, are they going to just ignore it and continue promoting the measure,

are they going to react to it and make some alteration, or you know, it's like what's the explanatory – what's the response.

Michael Lieberman: Yes.

Rute Martins: This is Rute. I agree that the measure developers should. When we are faced with low feasibility score, we need to assess what we're going to do next, whether that is opting for another data element that could be (inaudible) or it was trying to capture modifying what we intended in the beginning to be more feasible or sticking with what we've had because it's so important and it's a gap in care, for instance, that really should be a data element that is documented and isn't, for instance.

But there should be at least some idea of how the measure developer is going to deal with it. Now, of course, implementing this and incorporating into the workflow is going to be out of the scope for the measure developer, but the recommendation on what to do next, and how we're going to deal with it within the scope of the measure, I think, is part of our ...

Michael Lieberman: Yes. So, Reva, I think I agree with you. I think that the – and Rute as well. But I think it's with plans for improving the score, so I think we need to reword that a little bit.

Reva Winkler: OK.

Zahid Butt: Mike, this is Zahid. I think that's right. I was just – that's what I was going to say that I think that if they come up with, you know, a data element that scores low, then obviously it's not their sort of, in a sense, job to improve that score, but I think, to the extent that when they have an overall measure feasibility, then it's obviously important if they include that within that, within the framework of the overall eMeasure specification.

If they are going to include, you know, number one elements, then they need to have a plan of they want to improve it. But if the data element level is determined that is number one and they don't want to use it anymore, then it shouldn't be their job to then improve that particular element, right? If I'm reading this correctly.

Michael Lieberman: Yes.

Zahid Butt: So, I think I got that clarified that, you know, at one level it's not their job, but if they include the number one within the specification, then they need to have a plan on how they're going to improve that.

Michael Lieberman: Or even – yes, I would say kind of they have to – they have to provide an explanation of, you know, what they're going to do, what they're going to do or not do about the score, but I'll leave it up to Reva to come up with better wording.

Zahid Butt: If number one elements are included in the specification.

Reva Winkler: Right. That's ...

Michael Lieberman: Yes. OK. Next comment is line 48. Isn't this – isn't this addressed elsewhere in the measure development process because I don't think this is really feasibility as well.

Zahid Butt: Do we apply anything ruthlessly in our report?

Michael Lieberman: Isn't this about – I mean, this kind of measure measurement in general and not feasibility?

Female: I think they may be speaking about to the idea that certain data elements will, if not feasible, will have a greater impact on the measure, and I guess they're making suggestions on what has to be feasible for the measure to be valid. So, I think this does go beyond feasibility.

Michael Lieberman: OK. Then, line 49. OK, we've already addressed the first part. So, the second question about, you know, settings of care and that sort of thing, do we – do we mention that anywhere? I mean, I think it would be worthwhile to add that as an inspirational comment when we talk about the various numbers of, you know, multiple EHR systems. And do we say, you know, multiple types of care or not?

Female: No, we have one that's specific.

Zahid Butt: I think it would depend upon the intent of the measures, though.

Michael Lieberman: Yes.

Zahid Butt: You know.

Michael Lieberman: Yes. And that's why I would keep it pretty broad, but you know, across the types of the various types of locations where a measure would be applicable.

Zahid Butt: I think we sort of when we discussed the implementations, we purposely backed off being prescriptive about anything within that, and we just said it preferably should be multiple implementations, or I'm not even sure if that got in. I know Reva mentioned that.

Michael Lieberman: We did – I think we said that the other question would be, when we talk about where it was assessed, didn't we say it should, you know, you should capture various amount of data about that site and that would be whether it should come in, you know, and I don't know what exactly that amount of data would be, but it would be a site for practice and that sort of thing.

Female: Well, I think also there's something about that this rule should apply no matter what the care venue that's being assessed.

Michael Lieberman: Right. Yes. Or across – yes, I was thinking of like across ideally would be evaluated at various locations, types of locations.

Rute Martins: So, I'm curious – this is Rute. I'm curious about the language that NQF currently have for the validity and reliability of measures, regardless of whether their eMeasures are paper-based measures because it does speak a lot to the sample and what kind of variable should be taken into account.

And as far as I can tell, I don't think NQF imposes any kind of sample size and sample characteristics for that piece, so I would guess the same would apply here and we should be at least as broad as we are for the validity and reliability piece.

Michael Lieberman: OK. Reva, do you want to – can you add in a little bit of language around across care settings?

Reva Winkler: Yes, I can.

Michael Lieberman: I mean – and make it a recommendation, not a requirement.

Reva Winkler: Sure.

Michael Lieberman: OK. Then I think on number 49, the rest of those, we've addressed. We've addressed the number of installations and the last one is about minimum of EHR system seems low. I mean, we actually – I mean, I think we've already addressed that, although it's unlikely to move and naturally get a lot more than that, that would be better, but we've talked about that.

Next one, line 50, American College of Rheumatology appreciates those events. Yes, I think that that (segment) by rewording that comment for a plan for improving the score, and I think that we will try to reword that. So, that's first paragraph, the next paragraph, yes, I mean extremely high bar, so I think we've addressed both of those issues already.

Reva Winkler: Mike, this is Reva. Because we're talking about NQF's requirements and the guidance for endorsement, there were couple other things that have come up that I really would like to get the top spot on. One of them is we noticed internally, nowhere does NQF require that the measures be submitted in HQMF. Is that a reasonable requirement that the TEP agree, I think so, disagree.

Female: So, Reva, does NQF require HQMF or not?

Reva Winkler: At this point, we have not explicitly required it.

Female: All right.

Reva Winkler: And should we.

Michael Lieberman: So, I would – I mean, I would look to the measure developers to respond whether that's a reasonable requirement. I mean, if the developers we have on

the call feel that it is – that it's reasonable, that would be kind of the lowest bar to go over, because I think we also – you also have to think about, you know, non-professional measure developers and whether or not it would (inaudible) being able to submit an eMeasure, which may, you know, maybe all right. But if the professional developers don't feel it's reasonable, we need to know about it.

Paul Tang: I have heard, I'm not an expert HQMF, I've heard there are some required field in HQMF that may not be relevant to a specific measure, but then, of course, that creates overhead. And actually I've heard this from Epic. Howard, are you familiar wit this at all.

Michael Lieberman: Howard is on.

Howard Bregman: Sorry, my phone was ringing at the same time you asked that. Probably you have to say that again, I'm sorry.

Paul Tang: I've heard about there was some concerns about HQMF that it has some mandatory fields in it, so that may or may not be relevant to a specific quality measure, and so the mandatory field causes some unnecessary overhead, but I'm not an expert to know that.

Howard Bregman: Yes, I did hear something like that, but I don't know the details. Actually, on this issue, I would think that we don't need to make a statement about it.

Paul Tang: There is a separate question from NQF.

Michael Lieberman: Yes, whether we – whether we should require HQMF – whether H – whether NQF should require HQMF, and I think it gets in the feasibility issue about if you – and so that's can you express the measure in HQMF, and is that basically a minimum feasibility – minimal feasibility requirement or the path.

Rute Martins: This is Rute, and as a measure developer, I don't think we're considering right now developing any eMeasures that are not represented in HQMF because it would be – I don't even know how we would develop specifications that wouldn't rely on HQMF in the measure offering tool. But I would also say

that this is also a question for implementers. If they're given a measure, an eMeasure that is not an HQMF, what is the barrier to implementation.

Michael Lieberman: Yes.

Christopher Millet: This is Chris. As part of the whole idea behind, you know, creating eMeasures in HQMF so that the measures themselves are standardized, so that the different implementers and their systems can theoretically be able – once they can handle HQMF, they can handle any measure that's encoded in HQMF. So, I know that's part of the whole promise behind why HQMF and eMeasures were developed in the first place.

Female: Right.

Christopher Millet: And opposed to using the PDF document to keep up with any measure.

Ginny Meadows: Yes. And this is Ginny.

Female: Actually ...

Ginny Meadows: I think that's correct, and I guess I would ask if it's not in HQMF, how can we be assured of knowing that then when we actually implement that measure and run it to our population, that it would be given accurate result if we're having to do the interpretation and the formatting ourselves, I would be worried about that.

Michael Lieberman: Exactly. All right. So, it sounds like everybody that spoke would agree that the HQMF should be required for eMeasures submission. Anybody disagree? OK.

Reva Winkler: OK. That's helpful

Michael Lieberman: Any other questions, Reva? You know, we're at the end of our time.

Reva Winkler: Right. I mean, we need to see if there's anybody on the line for any public comment. I guess, the one thing I would ask the committee is the thought that feasibility as an NQF evaluation criterion is not a must have. And so, again, I

think Mike's earlier comment that said, it's not a matter of it being a threshold or an infeasible versus feasible, but some kind of high, moderate and low.

And I'm wondering, based on the kinds of things we're expecting people to assess, is there some way we could have a descriptor of what a high rating on feasibility would look like a moderate rating on feasibility and a low rating, and then insufficient information. Would that be useful to help address some of those issues?

Michael Lieberman: Reva, I would think that we would need – I mean it's really – that would be difficult to come up in a vacuum. I mean, I think we would need some experience with what measures look like and what they score in our experience with the measures to kind of come up with what those – what each one might look like.

Reva Winkler: OK. It brighten things up. OK. Well, since we're at the end, are there any public comment? Is anybody on the line who want to make a comment?

Alyssa Crawford: Hi. This is Alyssa Crawford from Mathematica. Can I make a really quick comment?

Reva Winkler: Sure.

Alyssa Crawford: So, regarding the last point that you discussed about the use of HQMF, and I have – I don't think we would have any objection to that. I just wanted to know that some things in measure logic can't be documented. In HQMF, it is kind of can't be documented through the math specifically.

And I'm thinking about composite, can't be entered into the math specifically, and I think there are still some discussion about how to best handle those, but I think it's something that needs to be recognized, because some measures may require either some hand modification or some supporting documentation of the specifications to explain things like that. So, I don't disagree with the use of HQMF, but I think some measures may need a little more than that to sort of peacefully document it.

Michael Lieberman: Thanks. OK.

Male: Could we say, at a minimum HQMF?

Rute Martins: This is Rute, and I actually think that's a really good point because if we require HQMF, are we saying that HQMF has everything that we need. And I think that's a really good point that if there are limitations to the HQMF, you may need more than the HQMF.

Reva Winkler: But you're OK with it being at least that as a minimum. And we can – we can put some language in there around the fact that there may be additional information needed in addition to the HQMF.

Michael Lieberman: Yes. I mean, you could – you could put in the language kind of around an exception policy. I mean, what you want to avoid is people just not doing HQMF because it's hard as opposed to not doing HQMF because you know, you want to do a composite measure and it doesn't support it.

Reva Winkler: Right.

Rute Martins: Right. And I agree with that.

Michael Lieberman: Yes.

Female: OK.

Michael Lieberman: All right.

Reva Winkler: Any other comments from anybody, public comments particularly?

OK, guys, since we're at the end of our time, we're kind of working on a short timeline to get this out. We want to be able to present it to HITAC and CSAC next week. So, the rest of this week, the team and I will be making the revision to the report that you guys have indicated.

We may need to come back actually with some questions by e-mail to clarify things, and we'd appreciate, you know, any responses you're able to provide for us. We will try and get some of the more major revisions out to you, you

know, just in a couple of days to be sure we've captured exactly what you think, but we are trying to get this finalized in the next couple of weeks.

Michael Lieberman: Great. And Reva, can you send that out as a marked up Word document or something so it's easy to find, that changes?

Reva Winkler: Yes. That's what we were planning on.

Michael Lieberman: OK. OK, great. OK. Well, thanks everybody for your time again, and we'll look forward to hearing from you, Reva.

Reva Winkler: OK, will do. Thanks everybody, really appreciate your time.

Michael Lieberman: Bye-bye.

END